

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 1, 2004, 17:21:01 ; Search time 10.2222 Seconds  
(without alignments)  
84.690 Million cell updates/sec

Title: US-09-905-083-33  
Perfect score: 40  
Sequence: 1 LLLPLQLLL 9

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 78: \*  
1: PIR1: \*  
2: PIR2: \*  
3: PIR3: \*  
4: PIR4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	40	100.0	253	2 A53968	serine proteinase
2	36	90.0	146	2 H75201	hypothetical prote
3	34	85.0	277	2 H84314	cytochrome aa3 con
4	34	85.0	369	2 S33603	surfactant protein
5	34	85.0	371	1 JN0450	conglutinin precu
6	34	85.0	371	2 I45878	conglutinin - bovi
7	34	85.0	754	2 A50614	probable competenc
8	34	85.0	845	2 D97163	cation transport p
9	33	82.5	196	2 G65039	hypothetical prote
10	33	82.5	282	2 E70890	hypothetical prote
11	33	82.5	370	2 AB3334	daunorubicin resis
12	33	82.5	398	2 C91063	hypothetical prote
13	33	82.5	413	2 AC0834	probable membrane
14	33	82.5	426	2 C93103	hypothetical prote
15	33	82.5	470	2 A90083	hypothetical prote
16	33	82.5	475	1 A59149	O-antigen transpor
17	33	82.5	653	2 A46362	amyloid precursor
18	32	80.0	198	2 S25656	T-cell surface gly
19	32	80.0	235	1 RWHU78	T-cell surface gly
20	32	80.0	238	1 LNR7WA	mannose-binding le
21	32	80.0	247	1 KYHUCM	chymase (EC 3.4.21
22	32	80.0	249	2 F91095	type III secretion
23	32	80.0	249	2 B85941	hypothetical prote
24	32	80.0	264	2 AC2620	1-acyl-sn-glycerol
25	32	80.0	264	2 C97402	probable acyltrans
26	32	80.0	266	2 JC7300	tax-responsive ele
27	32	80.0	267	2 JC4857	hepatocarcinogenes
28	32	80.0	491	2 JC6197	stromelysin 3 (EC
29	32	80.0	492	2 A44399	stromelysin 3 (EC

30	32	80.0	543	2 A12088	Na+/H+-exchanging
31	32	80.0	1306	1 A31759	peptidyl-dipeptida
32	31	77.5	100	2 A38685	apolipoprotein C-I
33	31	77.5	155	2 A31278	interleukin-2 prec
34	31	77.5	159	2 G75555	conserved hypothet
35	31	77.5	184	2 S10125	alpha-2u-globulin
36	31	77.5	189	2 T43766	hypothetical prote
37	31	77.5	244	2 F69260	nitrate ABC transp
38	31	77.5	249	2 T35589	probable secreted
39	31	77.5	269	2 T38931	hypothetical prote
40	31	77.5	303	2 C84914	hypothetical prote
41	31	77.5	372	2 A98157	probable permease
42	31	77.5	372	2 AH3130	ABC transporter, m
43	31	77.5	392	2 D83934	hypothetical prote
44	31	77.5	447	2 C94306	hypothetical prote
45	31	77.5	493	2 A71875	hypothetical prote

ALIGNMENTS

RESULT 1

A53968 serine proteinase SCCE precursor - human  
N:Alternate names: stratum corneum chymotryptic enzyme  
C:Species: Homo sapiens (man)  
C:Date: 07-Jul-1995 #sequence\_revision 07-Jul-1995 #text\_change 22-Jun-1999  
C:Accession: A53968  
R:Hansson, L.; Stroemqvist, M.; Baeckman, A.; Wallbrandt, P.; Carlstein, A.; Egelrud, T.  
J. Biol. Chem. 269 19420-19426, 1994  
A:Title: Cloning, expression, and characterization of stratum corneum chymotryptic enzyme  
A:Reference number: A53968; MUID:94308225; PMID:8034709  
A:Accession: A53968  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-253 <HAN>  
A:Cross-references: GB:L33404; NID:G521214; PIDN:AAC37551.1; PID:G532504  
C:Genetics:  
A:Gene: GDB:PRSS6; SCCE  
A:Cross-references: GDB:377730  
A:Map position: 7q35-7q35  
C:Superfamily: trypsin; trypsin homology  
F:30-245/Domain: trypsin homology <TRY>

Query Match 100.0%; Score 40; DB 2; Length 253;  
Best Local Similarity 100.0%; Pred. No. 1.8;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLLPLQLLL 9  
DB 5 LLLPLQLLL 13

RESULT 2

H75201 hypothetical protein PAB0088 - Pyrococcus abyssi (strain Orsay)  
C:Species: Pyrococcus abyssi  
C:Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 08-Sep-2000  
C:Accession: H75201  
R:anonymous, Genoscope  
submitted to the EMBL Data Library, July 1999  
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome struc  
A:Reference number: A75001  
A:Accession: H75201  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-146 <KAW>  
A:Cross-references: GB:AJ248283; GB:AL096836; NID:G5457433; PIDN:CAB49063.1; PID:el51495;  
A:Experimental source: strain Orsay  
C:Genetics:  
A:Gene: PAB0088  
C:Superfamily: Pyrococcus abyssi hypothetical protein PAB0088

```

Query Match      90.0%; Score 36; DB 2; Length 146;
Best Local Similarity 77.8%; Pred. No. 6.5;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLLPLQLLL 9
    |||||:::
Db 67 LLLPLQLIII 75

RESULT 3
cytochrome a3 controlling protein [imported] - Halobacterium sp. NRC-1
C:Species: Halobacterium sp. NRC-1
C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: H84314
R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.;
Leithausen, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jabl
Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li
A:Title: Genome sequence of Halobacterium species NRC-1.
A:Reference number: A84160; MUID:20504483; PMID:11016950
A:Accession: H84314
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-277 <STO>
A:Cross-references: GB:AE004437; NID:g10581096; PIDN:AAG19884.1; GSPDB:GN00138
C:Genetics:
A:Gene: ccp

Query Match      85.0%; Score 34; DB 2; Length 277;
Best Local Similarity 66.7%; Pred. No. 31;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLLPLQLLL 9
    :|||:::
Db 97 VLLPLQVIL 105

RESULT 4
surfactant protein D - bovine
C:Species: Bos primigenius taurus (cattle)
C>Date: 02-Dec-1993 #sequence_revision 01-Sep-1995 #text_change 17-Mar-1999
C:Accession: S33603
R:Dim, B.L.; Lu, J.; Reid, K.B.M.
Immunology 78, 159-165, 1993
A:Title: Structural similarity between bovine conglutinin and bovine lung surfactant pro
A:Reference number: S33603; MUID:93170856; PMID:8436402
A:Accession: S33603
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-369 <LIM>
C:Superfamily: pulmonary surfactant protein D; C-type lectin homology
F:248-367/Domain: C-type lectin homology <LCH>

Query Match      85.0%; Score 34; DB 2; Length 369;
Best Local Similarity 77.8%; Pred. No. 41;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQLLL 9
    |||||:::
Db 2 LLLPLSVLL 10

RESULT 5
JN0450
conglutinin precursor - bovine
N:Alternate names: C3b-binding protein
N:Contains: conglutinin-N
C:Species: Bos primigenius taurus (cattle)
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
C:Accession: JN0450; J02396; A23740; S33679; S35044; I46010; A29416; S34054

```

```

R:Suzuki, Y.; Yin, Y.P.; Makino, M.; Kurimura, T.; Wakamiya, N.
Biochem. Biophys. Res. Commun. 191, 335-342, 1993
A:Title: Cloning and sequencing of a cDNA coding for bovine conglutinin.
A:Reference number: JN0450; MUID:93213261; PMID:8460993
A:Accession: JN0450
A:Molecule type: mRNA
A:Residues: 1-371 <SUZ>
A:Cross-references: DBJ:D14085; NID:G285643; PIDN:BAA03170.1; PID:G285644
R:Kawasaki, N.; Itoh, N.; Kawasaki, T.
Biochem. Biophys. Res. Commun. 198, 597-604, 1994
A:Title: Gene organization and 5'-flanking region sequence of conglutinin: A C-type mamme
A:Reference number: J02396; MUID:94128104; PMID:8297370
A:Accession: J02396
A:Molecule type: mRNA
A:Residues: 1-371 <KA2>
A:Note: The authors translated the codon GAT for residues 250 and 270 as Glu
R:Lu, J.; Laursen, S.B.; Thiel, S.; Jensenius, J.C.; Reid, K.B.M.
Biochem. J. 292, 157-162, 1993
A:Title: The cDNA cloning of conglutinin and identification of liver as a primary site of
A:Reference number: S33235; MUID:93277452; PMID:7684896
A:Accession: S33235
A:Molecule type: mRNA
A:Residues: 1-172, 'H', 174-217, 'A', 219-271, 'V', 273-371 <LUJ>
A:Cross-references: EMBL:X71774; NID:G395267; PIDN:CAA50665.1; PID:G395268
A:Experimental source: liver
R:Lee, Y.M.; Leiby, K.R.; Allar, J.; Paris, K.; Lerch, B.; Okarma, T.B.
J. Biol. Chem. 266, 2715-2723, 1991
A:Title: Primary structure of bovine conglutinin, a member of the C-type animal lectin fa
A:Reference number: A23740; MUID:91131556; PMID:1993651
A:Accession: A23740
A:Molecule type: protein
A:Residues: 21-209, 'S', 211-371 <LEE>
R:Kawasaki, N.; Yokota, Y.; Kawasaki, T.
Arch. Biochem. Biophys. 305, 533-540, 1993
A:Title: Differentiation of conglutination activity and sugar-binding activity of conglut
A:Reference number: S36879; MUID:93384312; PMID:8373191
A:Accession: S36879
A:Molecule type: protein
A:Residues: 21-54, 75-86, 'X', 88-89, 'X', 91, 'X', 93-94, 208-209, 'X', 211-227 <KAW>
R:Lu, J.; Wiedemann, H.; Holmikov, U.; Thiel, S.; Timpl, R.; Reid, K.B.M.
Eur. J. Biochem. 215, 793-799, 1993
A:Title: Structural similarity between lung surfactant protein D and conglutinin. Two di
A:Reference number: S35044; MUID:93358905; PMID:8354286
A:Accession: S35044
A:Molecule type: protein
A:Residues: 75-86, 'X', 88-89, 'X', 91, 'I' <LUA>
R:Young, N.M.; Leon, M.A.
Biochem. Biophys. Res. Commun. 143, 645-651, 1987
A:Title: The carbohydrate specificity of conglutinin and its homology to proteins in the
A:Reference number: A29416; MUID:87184551; PMID:3566740
A:Contents: annotation
R:Malhotra, R.; Laursen, S.B.; Wallis, A.C.; Sim, R.B.
Biochem. J. 293, 15-19, 1993
A:Title: Research Communication. Localization of the receptor-binding site in the collect
A:Reference number: S34054; MUID:93319501; PMID:8328957
A:Contents: annotation
R:Liou, L.S.; Sastry, R.; Hartshorn, K.L.; Lee, Y.M.; Okarma, T.B.; Tauber, A.I.; Sastry,
J. Immunol. 153, 173-180, 1994
A:Title: Bovine conglutinin gene exon structure reveals its evolutionary relationship to
A:Reference number: I46010; MUID:94267222; PMID:8207234
A:Accession: I46010
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-371 <LIO>
A:Cross-references: EMBL:U06860; NID:G507183; PIDN:AAB60624.1; PID:G514256
C:Comment: This protein mediates the agglutination of erythrocytes with antibody and com
C:Comment: This protein is a Ca2+-dependent serum lectin specific for N-acetylglucosamine
C:Genetics:
A:Gene: CGN1
A:Introns: 67/1; 106/1; 142/1; 178/1; 217/1; 245/1

```

C:Superfamily: pulmonary surfactant protein D; C-type lectin homology  
 C:Keywords: calcium binding; glycoprotein; hydroxylysine; hydroxyproline  
 F:1-20/Domain: signal sequence #status predicted <SIG>  
 F:21-371/Product: conglutinin #status predicted <MAT>  
 F:46-214/Region: collagen-like  
 F:75-371/Product: conglutinin-N #status predicted <MA2>  
 F:248-369/Domain: C-type lectin homology <LCH>  
 F:63,87,99,135,141,159,162,198,210/Binding site: carbohydrate (Lys) (covalent) #status ex  
 F:63,87,99,135,141,159,162,198,210/Modified site: 5-hydroxylysine (Lys) #status experime  
 F:78,96,108,111,129,132,147,153,171,195/Modified site: 4-hydroxyproline (Pro) #status ex  
 F:337/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 85.0%; Score 34; DB 1; Length 371;  
 Best Local Similarity 77.8%; Pred. No. 42;  
 Matches 7; Conservative 1; Mismatches 0; Gaps 0;

QY 1 LLLPLQILL 9  
 |||||:  
 DB 2 LLLPLSVLL 10

## RESULT 6

I45878  
 conglutinin - bovine  
 C:Species: Bos primigenius taurus (cattle)  
 C:Date: 19-Dec-1997 #sequence\_revision 19-Dec-1997 #text\_change 20-Aug-1999  
 C:Accession: I45878  
 R:Liou, L.S.; Sastry, R.; Hartshorn, K.L.; Lee, Y.M.; Okarma, T.B.; Tauber, A.I.; Sastry  
 Gene 141, 277-281, 1994  
 A:Title: Bovine conglutinin (BC) mRNA expressed in liver: cloning and characterization o  
 A:Reference number: I45878; MUID:94215917; PMID:8163202  
 A:Accession: I45878  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA

A:Residues: 1-371 <LLO>  
 A:Cross-references: GB:L18871; NID:9495012; PIDN:AAA20126.1; PID:9495013  
 C:Superfamily: pulmonary surfactant protein D; C-type lectin homology  
 F:248-369/Domain: C-type lectin homology <LCH>

Query Match 85.0%; Score 34; DB 2; Length 371;  
 Best Local Similarity 77.8%; Pred. No. 42;  
 Matches 7; Conservative 1; Mismatches 0; Gaps 0;

QY 1 LLLPLQILL 9  
 |||||:  
 DB 2 LLLPLSVLL 10

## RESULT 7

AE0614  
 probable competence-related protein STY0984 [imported] - Salmonella enterica subsp. ente  
 C:Species: Salmonella enterica subsp. enterica serovar Typhi  
 A:Note: this species has also been called Salmonella typhi  
 C:Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 25-Aug-2003  
 C:Accession: AE0614  
 R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,  
 th, T.; Comerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,  
 S.; Moule, S.; O'Gaora, P.  
 Nature 413, 848-852, 2001  
 A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.  
 A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov  
 A:Reference number: AB0502; MUID:21534947; PMID:11677608  
 A:Accession: AE0614  
 A:Status: preliminary  
 A:Molecule type: DNA

A:Residues: 1-754 <PAR>  
 A:Cross-references: GB:AL513382; PIDN:CAD05383.1; PID:g16502146; GSPDB:GN00176  
 C:Genetics:

A:Gene: STY0984

C:Superfamily: competence protein ComEC

Query Match 85.0%; Score 34; DB 2; Length 754;  
 Best Local Similarity 77.8%; Pred. No. 86;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 LLLPLQILL 9  
 |||||:  
 DB 371 LLLPLQVAL 379

## RESULT 8

D97163  
 cation transport P-type ATPase CAC2137 [imported] - Clostridium acetobutylicum  
 C:Species: Clostridium acetobutylicum  
 C:Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 14-Sep-2001  
 C:Accession: D97163  
 R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,  
 J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.

J. Bacteriol. 183, 4823-4838, 2001

A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clos

A:Reference number: A96900; MUID:21359325; PMID:21359325

A:Accession: D97163

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-845 <KUR>

A:Cross-references: GB:AE001437; PIDN:AAK80095.1; PID:g15025128; GSPDB:GN00168

A:Experimental source: Clostridium acetobutylicum ATCC824

C:Genetics:

A:Gene: CAC2137

Query Match 85.0%; Score 34; DB 2; Length 845;  
 Best Local Similarity 87.5%; Pred. No. 97;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 LLLPLQILL 9  
 |||||:  
 DB 670 LLLPLQILL 677

## RESULT 9

G65039

hypothetical protein b2612 - Escherichia coli (strain K-12)

C:Species: Escherichia coli

C:Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 01-Mar-2002

C:Accession: G65039

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Col

A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97428617; PMID:9278503

A:Accession: G65039

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-196 <BLAT>

A:Cross-references: GB:AE000347; GB:U00096; NID:g2367142; PIDN:AACT5661.1; PID:g1788965;

A:Experimental source: strain K-12, substrain MGL655

Query Match 82.5%; Score 33; DB 2; Length 196;  
 Best Local Similarity 77.8%; Pred. No. 34;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9  
 |||||:  
 DB 123 LLLPLQILL 131

## RESULT 10

E70890

hypothetical protein Rv1978 - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999

C:Accession: E70890

R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.

; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;

Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A;Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
 A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
 A;Reference number: A70500; PMID:98295987; PMID:9634230

A;Accession: E70890  
 A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A;Molecule type: DNA

A;Residues: 1-282 <COL>

A;Cross-references: GB:AL022073; GB:AL123456; NID:93256024; PIDN:CAA17851.1; PID:e125651  
 A;Experimental source: strain H37RV

C;Genetics:  
 A;Gene: RV1978

Query Match 82.5%; Score 33; DB 2; Length 282;  
 Best Local Similarity 77.8%; Pred. No. 50;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9

Db 196 LLLPLHLL 204

RESULT 11  
 AB3334  
 daunorubicin resistance transmembrane protein [imported] - Brucella melitensis (strain 1

C;Species: Brucella melitensis  
 C;Date: 01-Feb-2002 #sequence\_revision 01-Feb-2002 #text\_change 01-Feb-2002

C;Accession: AB3334  
 R;DelVecchio, V.G.; Kaputal, V.; Redkar, R.J.; Patra, G.; Mujar, C.; Los, T.; Ivanova,

Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002

A;Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis

A;Reference number: AD3252; PMID:11756688

A;Accession: AB3334  
 A;Status: preliminary

A;Molecule type: DNA  
 A;Residues: 1-370 <KUR>

A;Cross-references: GB:AEQ08917; PIDN:AAL51837.1; PID:G17982584; GSPDB:GN00190

A;Experimental source: strain 16M

C;Genetics:

A;Gene: BME10656

A;Map position: 1

Query Match 82.5%; Score 33; DB 2; Length 370;

Best Local Similarity 75.0%; Pred. No. 65;

Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLLPLQIL 8

Db 293 LLLPLQVL 300

RESULT 12  
 C91063  
 hypothetical protein ECs3475 [imported] - Escherichia coli (strain O157:H7, substrain R1

C;Species: Escherichia coli  
 C;Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 24-Aug-2001

C;Accession: C91063  
 R;Hayaishi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.

Gasawara, N.; Yamanaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001

A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene

A;Reference number: A99629; PMID:21156231; PMID:11258796

A;Accession: C91063  
 A;Status: preliminary

A;Molecule type: DNA  
 A;Residues: 1-398 <HAY>

A;Cross-references: GB:BA000007; PIDN:BA36898.1; PID:G13362946; GSPDB:GN00154

A;Experimental source: strain O157:H7, substrain R1MD 050952

C;Genetics:

A;Gene: ECs3475

C;Superfamily: hypothetical protein HI0107

Query Match 82.5%; Score 33; DB 2; Length 398;

Best Local Similarity 77.8%; Pred. No. 71;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9

Db 101 LLLPLQILL 109

RESULT 13

AC0834

probable membrane protein corB [imported] - Salmonella enterica subsp. enterica serovar Typhi

C;Species: Salmonella enterica subsp. enterica serovar Typhi  
 A;Note: This species has also been called Salmonella typhi

C;Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002

C;Accession: AC0834

R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,

th, T.; Conner, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,

, S.; Moule, S.; O'Gaora, P.

Nature 413, 848-852, 2001

A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;

A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar

A;Reference number: AB0502; PMID:21534947; PMID:11677608

A;Accession: AC0834

A;Status: preliminary

A;Molecule type: DNA

A;Cross-references: GB:AL513382; PIDN:CAD05858.1; PID:G16503833; GSPDB:GN00176

C;Genetics:  
 A;Gene: corB

C;Superfamily: hypothetical protein HI0107

Query Match 82.5%; Score 33; DB 2; Length 413;

Best Local Similarity 77.8%; Pred. No. 73;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9

Db 116 LLLPLQILL 124

RESULT 14

C83103

hypothetical protein PA4338 [imported] - Pseudomonas aeruginosa (strain PA01)

C;Species: Pseudomonas aeruginosa  
 C;Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000

C;Accession: C83103

R;Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Br

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,

; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho

A;Reference number: AB2950; PMID:20437337; PMID:10984043

A;Accession: C83103

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-426 <STO>

A;Cross-references: GB:AE004850; GB:AE004091; NID:G9950560; PIDN:AA07726.1; GSPDB:GN001

A;Experimental source: strain PA01

C;Genetics:

A;Gene: PA4338

Query Match 82.5%; Score 33; DB 2; Length 426;

Best Local Similarity 77.8%; Pred. No. 76;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9

Db 301 LLLPLQILL 309

RESULT 15

AG0083

hypothetical protein orf470 [imported] - Guillardia theta nucleomorph



Wed Mar 3 07:49:10 2004

C:Species: nucleomorph Guillardia theta  
 A:Note: a nucleomorph is the vestigial nucleus of a eukaryotic endosymbiont  
 C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 24-May-2001  
 C:Accession: A90083  
 R:Douglas, S.; Zauner, S.; Fraunholz, M.; Beaton, M.; Penny, S.; Deng, L.T.; Wu, X.; Re  
 Nature 410, 1091-1096, 2001  
 A:Title: The highly reduced genome of an enslaved algal nucleus.  
 A:Reference number: A99082; MUID:11323671; PMID:11323671  
 A:Accession: A90083  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-470 <DOU>  
 A:Cross-references: GB:AF165818; NID:G6690603; PIDN:AAF24211.1; GSPDB:GN00150  
 C:Genetics:  
 A:Gene: orf470  
 A:Map position: 1  
 A:Genome: nucleomorph  
 C:Keywords: nucleomorph

Query Match 82.5%; Score 33; DB 2; Length 470;  
 Best Local Similarity 66.7%; Pred. No. 84;  
 Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLLPQLLL 9  
 :|||:|:  
 Db 253 MLLPLRLI 261

Search completed: March 1, 2004, 17:36:28  
 Job time : 12.3333 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 1, 2004, 17:17:25 ; Search time 6 Seconds  
(without alignments)  
78.105 Million cell updates/sec

Title: US-09-905-083-33  
Perfect score: 40  
Sequence: 1 LLLPLQLL 9

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141581 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141581

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	40	100.0	253	KLK7_HUMAN	P49862 homo sapien
2	38	95.0	201	FKBP_MOUSE	Q9d1m7 mus musculus
3	38	95.0	361	IHA_TRIUVI	O77755 trichosurus
4	34	85.0	369	PSPD_BOVIN	P35246 bos taurus
5	34	85.0	371	CL46_BOVIN	Q8mhz9 bos taurus
6	34	85.0	371	CONG_BOVIN	P23805 bos taurus
7	33	82.5	428	VFJD_ECOLI	P37908 escherichia
8	33	82.5	435	YA06_HUMAN	O60813 homo sapien
9	33	82.5	500	YA04_HUMAN	O60810 homo sapien
10	33	82.5	653	APPL_MOUSE	Q03157 mus musculus
11	32	80.0	198	CD8A_PONPY	P30433 pongo pygma
12	32	80.0	235	CD8A_HUMAN	P01732 homo sapien
13	32	80.0	238	NABA_RAT	P19999 rattus norv
14	32	80.0	239	CD8A_CANFA	P23706 canis fami
15	32	80.0	247	MCT1_HUMAN	P23946 homo sapien
16	32	80.0	247	MCT1_MACFA	P56435 macaca fasc
17	32	80.0	247	MCT1_PAPHA	P52195 papio hamad
18	32	80.0	492	MM11_MOUSE	Q02833 mus musculus
19	32	80.0	1306	ACE_HUMAN	P28212 homo sapien
20	31	77.5	100	APC2_CAVPO	P27916 cavia porce
21	31	77.5	153	IL2_RABIT	O77620 oryctolagus
22	31	77.5	155	IL2_RAT	P17108 rattus norv
23	31	77.5	184	MUP3_MOUSE	P04939 mus musculus
24	31	77.5	245	CS21_MOUSE	Q9ime3 mus musculus
25	31	77.5	246	CS21_HUMAN	Q9ime3 mus musculus
26	31	77.5	249	YBG0_STRCO	P40179 streptomyce
27	31	77.5	338	YDJ4_SCHPO	P87052 schizosacch
28	31	77.5	461	NCB1_HUMAN	Q02818 homo sapien
29	31	77.5	500	C72Z_ARATH	Q91xm3 arabidopsis
30	31	77.5	592	S132_HUMAN	Q13183 homo sapien
31	31	77.5	625	TR11_MOUSE	O35305 mus musculus
32	31	77.5	629	ACH4_MOUSE	P00174 mus musculus
33	31	77.5	630	ACH4_RAT	P09483 rattus norv

ALIGNMENTS

RESULT 1

ID	KLK7_HUMAN	STANDARD;	PRT;	253 AA.
AC	P49862;			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Kallikrein 7 precursor (EC 3.4.21.-) (Stratum corneum chymotryptic enzyme) (hSCCE).			
GN	KLK7 OR PRSS6 OR SCCE.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_taxid=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 23-53.			
RC	TISSUE=Skin;			
RX	MEDLINE=94308225; PubMed=8034709;			
RA	Hansson L., Stroemqvist M., Baeckman A., Wallbrandt P., Carlstein A.,			
RA	Egelrud T.,			
RT	"Cloning, expression, and characterization of stratum corneum			
RT	chymotryptic enzyme. A skin-specific human serine proteinase.";			
RL	J. Biol. Chem. 269:19420-19428(1994).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Keratinocytes;			
RA	Yousef G.M., Scorialas A., Diamandis E.P.;			
RT	"Molecular characterization, mapping and tissue expression of the			
RT	human stratum corneum chymotryptic enzyme gene.";			
RL	Submitted (DEC-1999) to the EMBL/GenBank/DBSJ databases.			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=20510030; PubMed=11054574;			
RA	Gan L., Lee I., Smith R., Argonza-Barrett R., Lei H., McCuaig J.,			
RA	Moss P., Paepser B., Wang K.;			
RT	"Sequencing and expression analysis of the serine protease gene			
RT	cluster located in chromosome 19q13 region.";			
RL	Gene 257:119-130(2000).			
RN	[4]			
RP	SEQUENCE FROM N.A.			
RA	Hansson L., Backman A., Ny A., Edlund M., Edholm E., Tornell J.,			
RA	Wallbrandt P., Egelrud T.;			
RT	"Epidermal overexpression of stratum corneum chymotryptic enzyme in			
RT	mice; a model for chronic itchy dermatitis.";			
RL	Submitted (DEC-2000) to the EMBL/GenBank/DBSJ databases.			
RN	[5]			
RP	CHARACTERIZATION.			
RX	MEDLINE=95314630; PubMed=7794273;			
RA	Skytt A., Stroemqvist M., Egelrud T.;			
RT	"Primary substrate specificity of recombinant human stratum corneum			
RT	chymotryptic enzyme.";			
RL	Biochem. Biophys. Res. Commun. 211:586-589(1995).			

-I- FUNCTION: May catalyze the degradation of intercellular cohesive structures in the cornified layer of the skin in the continuous shedding of cells from the skin surface. Specific for amino acid residues with aromatic side chains in the P1 position. SCCE

P51693 homo sapien  
P49446 mus musculus  
P23469 homo sapien  
Q9d866 mus musculus  
Q9y514 homo sapien  
P97798 mus musculus  
Q967q0 homo sapien  
P98161 homo sapien  
P43030 sus scrofa  
P26994 pseudomonas  
P26891 sus scrofa  
P05016 bos taurus

CC cleaves insulin B chain at 6-leu-|-Cys-7, 16-Tyr-|-Leu-17, 25-Phe-  
 CC |Tyr-26, and 26-Tyr-|-Thr-27. Could play a role in the activation  
 CC of precursors to inflammatory cytokines.  
 CC -!- TISSUE SPECIFICITY: Abundantly expressed in the skin and is  
 CC also seen in the brain and kidney.  
 CC -!- SIMILARITY: Belongs to peptidase family S1. Kallikrein subfamily.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; L33404; AAC37551.1; -  
 CC EMBL; AF166330; AAD49718.1; -  
 CC EMBL; AF243527; AAG33360.1; -  
 CC EMBL; AF332583; AAK69624.1; -  
 CC FIR; A53968; A53968.  
 CC HSP; P00763; IDPO.  
 CC MEROPS; S01.300; -  
 CC Genew; HGNC:6368; KLK7.  
 CC MIM; 604438; -  
 CC GO; GO:0008236; F-serine-type peptidase activity; TAS.  
 CC GO; GO:0008544; Epidermal differentiation; TAS.  
 CC InterPro; IPR009003; Cys\_Ser\_trypsin.  
 CC InterPro; IPR001254; Peptidase\_S1.  
 CC InterPro; IPR001314; Peptidase\_S1A.  
 CC Pfam; PF00089; trypsin; 1.  
 CC PRINTS; PR00722; CHYMOTRYPSIN.  
 CC SMART; SM00020; TRYP\_SPC; 1.  
 CC PROSITE; PS00240; TRYP\_SIN\_DOM; 1.  
 CC PROSITE; PS00134; TRYP\_SIN\_HIS; 1.  
 CC PROSITE; PS00135; TRYP\_SIN\_SER; 1.  
 CC Hydrolase; Serine protease; Zymogen; Glycoprotein; Signal.  
 FT SIGNAL 1 22  
 FT PROPEP 23 29  
 FT CHAIN 30 253  
 FT ACT\_SITE 70 70  
 FT ACT\_SITE 112 112  
 FT ACT\_SITE 205 205  
 FT ACT\_SITE 205 205  
 FT DISULFID 36 137  
 FT DISULFID 55 71  
 FT DISULFID 144 211  
 FT DISULFID 176 190  
 FT DISULFID 201 226  
 FT CARBOHYD 246 246  
 SQ SEQUENCE 253 AA; 27525 MW; 2D68B6B15A76A668 CRC64;  
 Query Match 100.0%; Score 40; DB 1; Length 253;  
 Best Local Similarity 100.0%; Pred. No. 0.97; Mismatches 0; Gaps 0;  
 Matches 9; Conservative 0; Indels 0; Indels 0;  
 QY 1 LLLPLQLL 9  
 Db 5 LLLPLQLL 13  
 RESULT 2  
 FKIL\_MOUSE STANDARD; PRT; 201 AA.  
 AC Q9D1N7; Q9CRE4;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE FK506 binding protein 11 precursor (EC 5.2.1.8) (Peptidyl-prolyl cis-  
 DE trans isomerase) (PPIase) (Notamase) (19 kDa FK506-binding protein)  
 DE (FKBP-19).  
 GN FKBP11.  
 OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Embryo;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 RA Aizawa K., Iwama M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,  
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
 RA Schriml L.M., Staabli F., Suzuki R., Tomika M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Rinaldo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,  
 RA Hayashizaki Y.;  
 RT "Functional annotation of a full-length mouse cDNA collection.";  
 RL Nature 409:685-690(2001).  
 RN (2)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Mammary gland;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Klausner R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Strausberg R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diachenko L., Marudina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stabile M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mallaby S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richardson D.K., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D., Muzly D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalley D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 CC -!- FUNCTION: PPIases accelerate the folding of proteins during  
 CC protein synthesis  
 CC -!- CATALYTIC ACTIVITY: Peptidylproline (omega=180) = peptidylproline  
 CC (omega=0).  
 CC -!- SIMILARITY: Belongs to the FKBP-type PPIase family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AK003331; BAB22719.1; -  
 CC EMBL; AK019132; BAB31559.1; -  
 CC EMBL; BC037596; AAK37596.1; -  
 CC HSP; P20071; 1FKJ  
 CC MGD; MGI:1913370; Fkbp11  
 CC InterPro; IPR001179; FKBP\_PPIase.  
 CC Pfam; PF00254; FKBP; 1.  
 CC PROSITE; PS00453; FKBP\_PPIASE\_1; 1.

DR PROSITE; PS00454; FKBP PPIASE 2; FALSE\_NEG.  
 DR PROSITE; PS00059; FKBP PPIASE\_3; 1.  
 KW Isomerase; Rotamase; Signal.  
 FT SIGNAL 1 POTENTIAL.  
 FT CHAIN 28 201 FKBP BINDING PROTEIN 11.  
 FT DOMAIN 57 144 PPIASE, FKBP-TYPE.  
 FT CONFLICT 53 53 S -> F (IN REF. 1; BAB31559).  
 FT CONFLICT 198 198 S -> R (IN REF. 1; BAB31559).  
 SQ SEQUENCE 201 AA; 22137 MW; 94D955C57264BD82 CRC64;  
 Query Match 95.0%; Score 38; DB 1; Length 201;  
 Best Local Similarity 88.9%; Pred. No. 19;  
 Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 LLLPLQLLL 9  
 Db 6 LLLPLQLLL 14  
 RESULT 3  
 ID IHA\_TRIVU STANDARD; PRT; 361 AA.  
 AC O07755;  
 DT 15-JUL-1999 (Rel. 38, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Inhibin alpha chain precursor.  
 GN INHA.  
 OS Trichosurus vulpecula (Brush-tailed possum).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Metatheria; Diprotodontia; Phalangeridae; Trichosurus.  
 RN NCBI\_TaxID=9337;  
 RX MEDLINE=99027340; PubMed=9801457;  
 RA Vamontford D., Fidler A.B., Heath D.A., Lawrence S.B., Tisdall D.J.,  
 RA Greenwood P.J., McNatty K.;  
 RT "cDNA sequence analysis, gene expression and protein localisation of  
 the inhibin alpha subunit of Australian brushtail possum (Trichosurus  
 vulpecula).";  
 RL J. Mol. Endocrinol. 21:141-152(1998).  
 CC -!- FUNCTION: INHIBINS AND ACTIVINS INHIBIT AND ACTIVATE,  
 RESPECTIVELY, THE SECRETION OF FOLLITROPIN BY THE PITUITARY GLAND.  
 INHIBINS/ACTIVINS ARE INVOLVED IN REGULATING A NUMBER OF DIVERSE  
 FUNCTIONS SUCH AS HYPOTHALAMIC AND PITUITARY HORMONE SECRETION,  
 GONADAL HORMONE SECRETION, GERM CELL DEVELOPMENT AND MATURATION,  
 ERYTHROID DIFFERENTIATION, INSULIN SECRETION, NERVE CELL SURVIVAL,  
 EMBRYONIC AXIAL DEVELOPMENT OR BONE GROWTH, DEPENDING ON THEIR  
 SUBUNIT COMPOSITION. INHIBINS APPEAR TO OPPOSE THE FUNCTIONS OF  
 ACTIVINS.  
 CC -!- SUBUNIT: DIMERIC, LINKED BY ONE OR MORE DISULFIDE BONDS.  
 INHIBIN A IS A DIMER OF ALPHA AND BETA-A.  
 INHIBIN B IS A DIMER OF ALPHA AND BETA-B.  
 CC -!- SIMILARITY: Belongs to the TGF-beta family.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 DR EMBL; AF033340; AAC63945.1;  
 DR GO; GO:0005576; C:extracellular; ISS.  
 DR GO; GO:0017106; F:activin inhibitor activity; ISS.  
 DR GO; GO:0005125; F:cytokine activity; ISS.  
 DR GO; GO:0003793; F:defense/immunity protein activity; ISS.  
 DR GO; GO:0008083; F:growth factor activity; ISS.  
 DR GO; GO:0005179; F:hormone activity; ISS.  
 DR GO; GO:0005515; F:protein binding; ISS.  
 DR GO; GO:0007030; P:cell cycle arrest; ISS.  
 DR GO; GO:00030154; P:cell differentiation; ISS.

DR GO; GO:0008151; P:cell growth and/or maintenance; ISS.  
 DR GO; GO:0007166; P:cell surface receptor linked signal transdu. .; ISS.  
 DR GO; GO:0007267; P:cell-cell signaling; ISS.  
 DR GO; GO:030218; P:erythrocyte differentiation; ISS.  
 DR GO; GO:042541; P:hemooglobin biosynthesis; ISS.  
 DR GO; GO:0008917; P:induction of apoptosis; ISS.  
 DR GO; GO:0045578; P:negative regulation of B-cell differentiation; ISS.  
 DR GO; GO:0045786; P:negative regulation of cell cycle; ISS.  
 DR GO; GO:0046882; P:negative regulation of follicle-stimulating. .; ISS.  
 DR GO; GO:0045077; P:negative regulation of interferon-gamma bio. .; ISS.  
 DR GO; GO:0045650; P:negative regulation of macrophage different. .; ISS.  
 DR GO; GO:0042326; P:negative regulation of phosphorylation; ISS.  
 DR GO; GO:0007399; P:neurogenesis; ISS.  
 DR GO; GO:0001541; P:ovarian follicle development; ISS.  
 DR GO; GO:0046881; P:positive regulation of follicle-stimulating. .; ISS.  
 DR InterPro; IPR002405; Inhibin\_alpha.  
 DR InterPro; IPR001839; TGFb.  
 DR Pfam; PF00019; TGF-beta; 1.  
 DR PRINTS; PRO0669; INHIBINA.  
 DR PRODOM; PD000357; TGFb; 1.  
 DR SMART; SM00204; TGFb; 1.  
 DR PROSITE; PS00250; TGF\_BETA\_1; 1.  
 KW Growth factor; Hormone; Glycoprotein; Signal.  
 FT SIGNAL 1 21 POTENTIAL.  
 FT PROPEP 22 230 BY SIMILARITY.  
 FT CHAIN 231 361 INHIBIN ALPHA CHAIN.  
 FT DISULFID 260 323 BY SIMILARITY.  
 FT DISULFID 289 358 BY SIMILARITY.  
 FT DISULFID 293 360 BY SIMILARITY.  
 FT DISULFID 322 322 INTERCHAIN (BY SIMILARITY).  
 FT CARBOHYD 48 144 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 144 144 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 266 266 N-LINKED (GLCNAC. .) (POTENTIAL).  
 SQ SEQUENCE 361 AA; 38945 MW; D661CDF93CDA87D CRC64;  
 Query Match 95.0%; Score 38; DB 1; Length 361;  
 Best Local Similarity 88.9%; Pred. No. 3.4;  
 Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 LLLPLQLLL 9  
 Db 4 LLLPLQLLL 12  
 RESULT 4  
 ID PSPD\_BOVIN STANDARD; PRT; 369 AA.  
 AC P35246;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Pulmonary surfactant-associated protein D precursor (SP-D) (PSP-D).  
 GN SFTPD OR SFTP4.  
 OS Bos taurus (Bovine).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 CC Bovidae; Bovinae; Bos.  
 CC -!- FUNCTION: Contributes to the lung's defense against inhaled  
 microorganisms. Binds strongly maltose residues and to a lesser  
 extent other alpha-glucosyl moieties. It could participate in the  
 extracellular reorganization or turnover of pulmonary surfactant.  
 CC -!- SUBUNIT: Oligomeric complex of 4 set of homotrimers.  
 CC -!- SUBCELLULAR LOCATION: Extracellular.  
 TISSUE=Lung;  
 MEDLINE=93170856; PubMed=8436402;  
 RA Lim B.L., Lu J., Reid K.B.M.;  
 RT "Structural similarity between bovine conglutinin and bovine lung  
 surfactant protein D and demonstration of liver as a site of  
 synthesis of conglutinin.";  
 RL Immunology 78:159-165(1993).  
 CC -!- FUNCTION: Contributes to the lung's defense against inhaled  
 microorganisms. Binds strongly maltose residues and to a lesser  
 extent other alpha-glucosyl moieties. It could participate in the  
 extracellular reorganization or turnover of pulmonary surfactant.  
 CC -!- SUBUNIT: Oligomeric complex of 4 set of homotrimers.  
 CC -!- SUBCELLULAR LOCATION: Extracellular.

CC -!- MISCELLANEOUS: Pulmonary surfactant consists of 90% lipid and 10%  
 CC protein. There are 4 surfactant-associated proteins: 2 collagenous,  
 CC carbohydrate-binding glycoproteins (SP-A and SP-D) and 2 small  
 CC hydrophobic proteins (SP-B and SP-C).  
 CC -!- SIMILARITY: Contains 1 collagenous domain.  
 CC -!- SIMILARITY: Contains 1 C-type lectin family domain.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: X75911; CAA53510.1; -  
 CC HSP: P35247; I808.  
 CC InterPro: IPR008160; Collagen.  
 CC InterPro: IPR001304; Lectin\_C.  
 CC Pfam: PF01391; Collagen; 2.  
 CC Pfam: PF00059; lectin\_c; 1.  
 CC SMART: SM00034; CLECT; 1.  
 CC PROSITE: PS00615; C-TYPE LECTIN 1; 1.  
 CC PROSITE: PS0041; C-TYPE LECTIN 2; 1.  
 CC Glycoprotein; Calcium; Surface film; Gaseous exchange; Hydroxylation;  
 CC Signal; Lectin; Collagen; Repeat; Coiled coil.  
 CC SIGNAL 1 20  
 CC FT CHAIN 21 369  
 CC  
 CC DOMAIN 46 216  
 CC FT DOMAIN 217 248  
 CC FT DOMAIN 273 369  
 CC FT DISULFID 275 367  
 CC FT DISULFID 345 359  
 CC FT CARBOHYD 90 90  
 CC FT MOD\_RES 78 78  
 CC FT MOD\_RES 87 87  
 CC FT MOD\_RES 96 96  
 CC FT MOD\_RES 99 99  
 CC FT MOD\_RES 165 165  
 CC FT MOD\_RES 171 171  
 CC SEQUENCE 369 AA; 37361 MW; 07D88B24E0ABE2E3 CRC64;  
 CC  
 CC Query Match 85.0%; Score 34; DB 1; Length 369;  
 CC Best Local Similarity 77.8%; Pred. No. 22;  
 CC Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 CC  
 CC QY 1 LLLPLQILL 9  
 CC |||||:|  
 CC Db 2 LLLPLSVLL 10  
 CC  
 CC RESULT 5  
 CC CL46\_BOVIN STANDARD; PRT; 371 AA.  
 CC AC Q8MH29;  
 CC DT 28-FEB-2003 (Rel. 41, Created)  
 CC DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 CC DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 CC DE Collectin-46 precursor (CL-46) (46 kDa collectin).  
 CC GN CL46.  
 CC OS Bos taurus (Bovine).  
 CC OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 CC OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 CC OC Bovidae; Bovinae; Bos.  
 CC OX NCBI\_TaxID=9913;  
 CC RN [1]  
 CC RP SEQUENCE FROM N.A.  
 CC RA Hansen S., Holm D., Moeller V., Vitved L., Bendixen C., Reid K.B.M.,  
 CC RA Skjold K., Holmskov U.;  
 CC RT "CL-46, a novel collectin highly expressed in the bovine thymus and  
 CC RT liver.";  
 CC RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.

CC -!- SUBUNIT: Oligomeric complex of 4 set of homotrimers (By  
 CC similarity).  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- TISSUE SPECIFICITY: Highly expressed in thymus and liver.  
 CC -!- PTM: Hydroxylated (Potential).  
 CC -!- SIMILARITY: Contains 1 collagenous domain.  
 CC -!- SIMILARITY: Contains 1 C-type lectin family domain.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: AF509589; AAM34742.1; -  
 CC EMBL: AF509590; AAM34743.1; -  
 CC InterPro: IPR008160; Collagen.  
 CC InterPro: IPR001304; Lectin\_C.  
 CC Pfam: PF01391; Collagen; 2.  
 CC Pfam: PF00059; lectin\_c; 1.  
 CC SMART: SM00034; CLECT; 1.  
 CC PROSITE: PS00615; C-TYPE LECTIN 1; 1.  
 CC PROSITE: PS0041; C-TYPE LECTIN 2; 1.  
 CC Lectin; Hydroxylation; Glycoprotein; Mannose-binding; Membrane;  
 CC Collagen; Repeat; Calcium; Signal.  
 CC SIGNAL 1 20  
 CC FT CHAIN 21 371  
 CC FT DOMAIN 46 216  
 CC FT DOMAIN 273 371  
 CC FT SITE 201 203  
 CC FT DISULFID 275 369  
 CC FT DISULFID 347 361  
 CC FT CARBOHYD 90 90  
 CC SEQUENCE 371 AA; 37445 MW; 108AC45A91420B93 CRC64;  
 CC  
 CC Query Match 85.0%; Score 34; DB 1; Length 371;  
 CC Best Local Similarity 77.8%; Pred. No. 22;  
 CC Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 CC  
 CC QY 1 LLLPLQILL 9  
 CC |||||:|  
 CC Db 2 LLLPLSVLL 10  
 CC  
 CC RESULT 6  
 CC CONG\_BOVIN STANDARD; PRT; 371 AA.  
 CC AC P23805; O97748;  
 CC DT 01-NOV-1991 (Rel. 20, Created)  
 CC DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 CC DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 CC DE Conglutinin precursor.  
 CC GN CGNI.  
 CC OS Bos taurus (Bovine).  
 CC OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 CC OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 CC OC Bovidae; Bovinae; Bos.  
 CC OX NCBI\_TaxID=9913;  
 CC RN [1]  
 CC RP SEQUENCE FROM N.A.  
 CC RA MEDLINE=93213261; PubMed=8460993;  
 CC RA Suzuki Y., Yin Y., Makino M., Kurimura T., Wakamiya N.;  
 CC RT "Cloning and sequencing of a cDNA coding for bovine conglutinin.";  
 CC RL Biochem. Biophys. Res. Commun. 191:335-342(1993).  
 CC  
 CC RN [2]  
 CC RP SEQUENCE FROM N.A.  
 CC RA MEDLINE=93277452; PubMed=7684896;  
 CC RA Lu J., Laursen S.B., Thiel S., Jensenius J.C., Reid B.M.;  
 CC RT "The cDNA cloning of conglutinin and identification of liver as a  
 CC RT primary site of synthesis of conglutinin in members of the Bovidae.";  
 CC RL Biochem. J. 292:157-162(1993).

RN RP SEQUENCE FROM N.A.  
 RC TISSUE=Liver;  
 RX MEDLINE=94215917; PubMed=8163202;  
 RA Liou L.S., Sastre R., Hartshorn K.L., Lee Y.M., Okarma T.B.,  
 RA Tauber A.I., Sastre K.N.; mRNA expressed in liver: cloning and  
 RT "Bovine conglutinin (BC) mRNA expressed in liver: cloning and  
 RT characterization of the BC cDNA reveals strong homology to surfactant  
 RL protein-D.";  
 RL Gene 141:277-281(1994).  
 [4]  
 RN RP SEQUENCE FROM N.A.  
 RC TISSUE=Semen;  
 RX MEDLINE=94267222; PubMed=8207234;  
 RA Liou L.S., Sastre R., Hartshorn K.L., Lee Y.M., Okarma T.B.,  
 RA Tauber A.I., Sastre K.N.;  
 RT "Bovine conglutinin gene exon structure reveals its evolutionary  
 RT relationship to surfactant protein-D.";  
 RL J. Immunol. 153:173-180(1994).  
 [5]  
 RN RP SEQUENCE FROM N.A.  
 RC TISSUE=Liver;  
 RX MEDLINE=94128104; PubMed=8297370;  
 RA Kawasaki N., Itoh N., Kawasaki T.;  
 RT "Gene organization and 5'-flanking region sequence of conglutinin: a  
 RT C-type mammalian lectin containing a collagen-like domain.";  
 RL Biochem. Biophys. Res. Commun. 198:597-604(1994).  
 [6]  
 RN RP SEQUENCE OF 21-371.  
 RX MEDLINE=91131556; PubMed=1993651;  
 RA Lee Y.-M., Leiby K.R., Allier J., Paris K., Lerch B., Okarma T.B.;  
 RT "Primary structure of bovine conglutinin, a member of the C-type  
 RT animal lectin family.";  
 RL J. Biol. Chem. 266:2715-2723(1991).  
 [7]  
 RN RP PRELIMINARY SEQUENCE OF 21-52.  
 RX MEDLINE=87184551; PubMed=3566740;  
 RA Young N.M., Leon M.A.;  
 RT "The carbohydrate specificity of conglutinin and its homology to  
 RT proteins in the hepatic lectin family.";  
 RL Biochem. Biophys. Res. Commun. 143:645-651(1987).  
 CC -!- FUNCTION: Calcium-dependent lectin-like protein which binds to a  
 CC yeast cell wall extract and immune complexes through the  
 CC complement component (C3bi). It is capable of binding nonreducing  
 CC terminal N-acetylglucosamine, mannose, and fucose residues.  
 CC -!- SUBUNIT: Oligomeric complex of 4 set of homotrimers.  
 CC -!- SIMILARITY: Contains 1 collagenous domain.  
 CC -!- SIMILARITY: Contains 1 C-type lectin family domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed, usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; D14085; BAA03170.1; -  
 DR EMBL; X71774; CAA50665.1; -  
 DR EMBL; U18871; AAB20126.1; -  
 DR EMBL; U06860; AAB60624.1; -  
 DR EMBL; U06854; AAB60624.1; JOINED.  
 DR EMBL; U06855; AAB60624.1; JOINED.  
 DR EMBL; U06856; AAB60624.1; JOINED.  
 DR EMBL; U06857; AAB60624.1; JOINED.  
 DR EMBL; U06858; AAB60624.1; JOINED.  
 DR EMBL; U06859; AAB60624.1; JOINED.  
 DR EMBL; D25302; BAA04983.2; -  
 DR EMBL; D25296; BAA04983.2; JOINED.  
 DR EMBL; D25297; BAA04983.2; JOINED.  
 DR EMBL; D25298; BAA04983.2; JOINED.  
 DR EMBL; D25299; BAA04983.2; JOINED.  
 DR EMBL; D25300; BAA04983.2; JOINED.

DR EMBL; D25301; BAA04983.2; JOINED.  
 DR PIR; I45878; I45878.  
 DR PIR; JN0450; JN0450.  
 DR HSSP; P35247; I308.  
 DR InterPro; IPR008161; Clg\_helix.  
 DR InterPro; IPR008160; Collagen.  
 DR InterPro; IPR001304; Lectin\_C.  
 DR Pfam; PF01391; Collagen; 3.  
 DR Pfam; PF00059; Lectin\_C; 1.  
 DR ProDom; PDC00007; Clg\_helix; 1.  
 DR SMART; SM00034; CLECT; 1.  
 DR PROSITE; PS00615; C-TYPE LECTIN 1; 1.  
 DR PROSITE; PS00615; C-TYPE LECTIN 2; 1.  
 KW Lectin; Hydroxylation; Glycoprotein; Mannose-binding; Membrane;  
 KW Collagen; Repeat; Calcium; Signal.  
 FT SIGNAL 1 20  
 FT CHAIN 21 371  
 FT DOMAIN 46 216  
 FT DOMAIN 273 371  
 FT MOD\_RES 63 63  
 FT MOD\_RES 87 87  
 FT MOD\_RES 99 99  
 FT MOD\_RES 135 135  
 FT MOD\_RES 141 141  
 FT MOD\_RES 159 159  
 FT MOD\_RES 162 162  
 FT MOD\_RES 198 198  
 FT SITE 201 203  
 FT DISULFID 275 369  
 FT DISULFID 337 361  
 FT CARBOHYD 337 337  
 FT CONFLICT 173 173  
 FT CONFLICT 210 210  
 FT CONFLICT 218 218  
 FT CONFLICT 272 272  
 SQ SEQUENCE 371 AA; 37994 MW; 867BB41992544B1F CRC64;  
 Query Match 85.0%; Score 34; DB 1; Length 371;  
 Best Local Similarity 77.8%; Pred. No. 22;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 LLLPLQLLL 9  
 DB 2 LLLPLSVLL 10  
 RESULT 7  
 YFJD\_ECOLI STANDARD; PRT; 428 AA.  
 ID YFJD\_ECOLI  
 AC P37908; P76600; P76601; P77009;  
 DT 01-OCT-1994 (Rel. 30, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical protein yfjd.  
 GN YFJD OR B2612/B2613.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=K12 / MG1655;  
 RC MEDLINE=97426617; PubMed=9278503;  
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12.";  
 RL Science 277:12453-1474(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=97349980; PubMed=9205837;

RA Yamamoto Y., Aiba H., Baba T., Hayashi K., Inada T., Isono K.,  
 RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,  
 RA Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nashimoto H.,  
 RA Oshima T., Oyama S., Saito N., Sampei G., Satoh Y., Sivasundaram S.,  
 RA Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,  
 RA Yamagata S., Horiuchi T.,  
 RT "Construction of a contiguous 874-kb sequence of the Escherichia coli  
 RT - K12 genome corresponding to 50.0-68.8 min on the linkage map and  
 RT analysis of its sequence features.";  
 RL DNA Res. 4:91-113(1997).  
 RN [3]  
 RP SEQUENCE OF 322-428 FROM N.A.  
 RP STRAIN-BJ78;  
 RC MEDLINE=88319942; PubMed=3045760;  
 EX Lipinska B., King J., Ang D., Georgopoulos C.;  
 RA "Sequence analysis and transcriptional regulation of the Escherichia  
 RT coli grpE gene, encoding a heat shock protein.";  
 RL Nucleic Acids Res. 16:7545-7562(1988).  
 RN [4]  
 RP IDENTIFICATION.  
 RA Rudd K.E.;  
 RL Unpublished observations (AUG-1994).  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
 CC -!- SIMILARITY: BELONGS TO THE UPF0053 FAMILY. STRONG, TO  
 CC H. INFLUENZAE HI10107.  
 CC -!- SIMILARITY: Contains 2 CBS domains.  
 CC -!- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A  
 CC FRAMESHIFT IN POSITION 199 THAT PRODUCES TWO ORFS.  
 CC -!- CAUTION: Ref.3 sequence differs from that shown due to numerous  
 CC frameshifts.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; AB000347; AAC75661.1; ALT FRAME.  
 DR EMBL; AE000347; AAC75662.1; ALT FRAME.  
 DR EMBL; D90888; BAA16497.1; ALT INIT.  
 DR EMBL; X07863; -; NOT ANNOTATED\_CDS.  
 DR EcoGene; EG12442; yfjD.  
 DR InterPro; IPR002550; CBS.  
 DR InterPro; IPR000644; CBS domain.  
 DR InterPro; IPR005170; CorC\_transp-asc.  
 DR Pfam; PF00571; CBS; 2.  
 DR Pfam; PF03471; CorC\_HlyC; 1.  
 DR Pfam; PF01595; DUF21; 1.  
 DR SMART; SM00116; CBS; 1.  
 DR Hypothetical protein; CBS domain; Repeat; Transmembrane;  
 KW Complete proteome.  
 FT TRANSMEM 4 24 POTENTIAL.  
 FT TRANSMEM 65 85 POTENTIAL.  
 FT TRANSMEM 92 112 POTENTIAL.  
 FT TRANSMEM 130 150 POTENTIAL.  
 SQ SEQUENCE 428 AA; 48044 MW; 38456865EDBCB151 CRC64;  
 Query Match 82.5%; Score 33; DB 1; Length 428;  
 Best Local Similarity 77.8%; Pred. No. 39;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 LLLPQILL 9  
 DB 131 LLAPQLILM 139  
 RESULT 8  
 YAO6 HUMAN  
 ID YAO6 HUMAN STANDARD; PRT; 435 AA.  
 AC O60813;  
 DT 30-MAY-2000 (Rel. 39, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Hypothetical protein DJ945024.5 (Fragment).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Grafham D.;  
 RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Rhodes S.;  
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
 CC -!- SIMILARITY: Belongs to the MAPE family.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; AL022101; CAAL7880.1; -;  
 DR EMBL; AL049680; CAB41252.1; -;  
 KW Hypothetical protein.  
 FT NON TER 435 435  
 SQ SEQUENCE 435 AA; 50539 MW; 3D19443032BBB494 CRC64;  
 Query Match 82.5%; Score 33; DB 1; Length 435;  
 Best Local Similarity 87.5%; Pred. No. 40;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 LLPLQILL 9  
 DB 298 LVPLQILL 305  
 RESULT 9  
 YAO4 HUMAN  
 ID YAO4 HUMAN STANDARD; PRT; 500 AA.  
 AC O60810;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Hypothetical protein DJ945024.2 (Fragment).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Grafham D.;  
 RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Rhodes S.;  
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
 CC -!- SIMILARITY: Belongs to the MAPE family.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; AL022101; CAAL7877.1; -;  
 DR EMBL; AL049681; CAB41253.1; -;  
 KW Hypothetical protein.

FT NON\_TER 1 500  
 TT SEQUENCE 500 AA; 58247 MW; 1855CD8A8F14B7C3 CRC64;  
 SQ

Query Match 82.5%; Score 33; DB 1; Length 500;  
 Best Local Similarity 87.5%; Pred. No. 46;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 LPLQLIILL 9  
 DB 363 LVPLQILL 370

RESULT 10  
 APPL MOUSE STANDARD; PRT; 653 AA.  
 AC Q03157; Q8VC38;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-OCT-1993 (Rel. 27, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Amyloid-like protein 1 precursor (APLP) (APLP-1) [Contains: C30].  
 GN APLP1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=93066322; PubMed=1279693;  
 RA Wasco W., Bupp K., Magendantz M., Gusella J.F., Tanzi R.E.,  
 RA Solomon F.,  
 RA "Identification of a mouse brain cDNA that encodes a protein related  
 RT to the Alzheimer disease-associated amyloid beta protein precursor";  
 RL Proc. Natl. Acad. Sci. U.S.A. 89:10758-10762(1992).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Retina;  
 RX MEDLINE=22398257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield V.S.N., Krzywinski M.I., Skalska U., Smalley D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 human and mouse cDNA sequences";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [3]  
 RP COLLAGEN-BINDING.  
 RX MEDLINE=96139497; PubMed=8576160;  
 RA Behr D., Hesse L., Masters C.L., Muthaup G.;  
 RT "Regulation of amyloid protein precursor (APP) binding to collagen and  
 RT mapping of the binding sites on APP and collagen type I";  
 RL J. Biol. Chem. 271:1613-1620(1996).  
 RN [4]  
 RP INTERACTION WITH DAB1.  
 RX MEDLINE=99389880; PubMed=10460257;  
 RA Homayouni R., Rice D.S., Sheldon M., Curran T.;  
 RT "Disabled-1 binds to the cytoplasmic domain of amyloid precursor-like  
 RT protein 1";  
 RL J. Neurosci. 19:7507-7515(1999).

[5]  
 RP INTERACTION WITH MAPK8IP1.  
 RX MEDLINE=21408156; PubMed=11517249;  
 RA Matsuda S., Yasukawa T., Homma Y., Ito Y., Niikura T., Hiraki T.,  
 RA Hirai S., Ohno S., Kita Y., Kawasumi M., Koyama K., Yamamoto T.,  
 RA Kyriakis J.M., Nishimoto I.;  
 RT "C-jun N-terminal kinase (JNK)-interacting protein-1b/Islet-brain-1  
 RT scaffolds Alzheimer's amyloid precursor protein with JNK";  
 RL J. Neurosci. 21:6597-6607(2001).  
 RN [6]  
 RP GAMMA-SECRETASE PROCESSING, INTERACTION WITH APBB1, AND MUTAGENESIS OF  
 RP TYR-641.  
 RX MEDLINE=22313598; PubMed=12228233;  
 RA Scheinfeld M.H., Ghersi E., Laky K., Powlkes B.J., D'Adamo L.;  
 RA "Processing of beta-amyloid precursor-like protein-1 and -2 by gamma-  
 RT secretase regulates transcription";  
 RL J. Biol. Chem. 277:44195-44201(2002).  
 CC -!- FUNCTION: May play a role in postsynaptic function. The C-terminal  
 CC gamma-secretase processed fragment, A1D1, activates transcription  
 CC activation through APBB1 (Fe65) binding. Couples to JIP signal  
 CC transduction through C-terminal binding. May interact with  
 CC cellular G-protein signaling pathways. Can regulate neurite  
 CC outgrowth through binding to components of the extracellular  
 CC matrix such as heparin and collagen I.  
 CC -!- FUNCTION: The gamma-CTF peptide, C30, is a potent enhancer of  
 CC neuronal apoptosis (By similarity).  
 CC -!- SUBUNIT: Binds, via its C-terminal, to the PID domain of several  
 CC cytoplasmic proteins, including APBB and APBA family members,  
 CC MAPK8IP1 and Dab1 (By similarity). Binding to Dab1 inhibits its  
 CC serine phosphorylation.  
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein. C-terminally  
 CC processed in the Golgi complex.  
 CC -!- DOMAIN: The NPXY sequence motif found in many tyrosine-  
 CC phosphorylated proteins is required for the specific binding of  
 CC the PID domain. However additional amino acids either N- or C-  
 CC terminal to the NPXY motif are often required for complete  
 CC interaction. The NPXY site is also involved in clathrin-mediated  
 CC endocytosis.  
 CC -!- PTM: Proteolytically cleaved by caspases during neuronal  
 CC apoptosis. Cleaved, in vitro, at Asp-623 by caspase-3 (By  
 CC similarity).  
 CC -!- PTM: N- and O-glycosylated.  
 CC -!- MISCELLANEOUS: Binds zinc and copper in the extracellular domain.  
 CC Zinc-binding increases heparin binding. No Cu(II) reducing  
 CC activity with copper-binding  
 CC -!- SIMILARITY: Belongs to the APP family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC -----  
 CC EMBL; L04538; AAA37247.1; ..  
 CC EMBL; BC021877; AAH21877.1; ..  
 CC PIR; A45362; A45362.  
 CC HSP; P05067; LMWP.  
 CC MGD; MGI:88046; Ap1p1.  
 CC InterPro; IPR008155; A4\_APP.  
 CC InterPro; IPR008154; A4\_extra.  
 CC Pfam; PF02177; A4\_EXTRA; 1.  
 CC PRINTS; PR00203; AMYLOIDA4.  
 CC SMART; SM00006; A4\_EXTRA; 1.  
 CC PROSITE; PS00319; A4\_EXTRA; 1.  
 CC PROSITE; PS00320; A4\_INTRA; 1.  
 CC PROSITE; PS00320; A4\_INTRA; 1.  
 CC KW Apoptosis; Endocytosis; Cell adhesion; Coated pits; Neurone;  
 CC Heparin-binding; Metal-binding; Copper; Zinc; Signal; Transmembrane;  
 CC Glycoprotein.  
 CC SIGNAL 1 37 POTENTIAL.  
 CC CHAIN 38 653 AMYLOID-LIKE PROTEIN 1.  
 CC CHAIN 624 653 C30 (BY SIMILARITY).  
 FT



FT DOMAIN 38 593 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 584 606 POTENTIAL.  
 FT DOMAIN 607 653 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 157 177 COPPER-BINDING.  
 FT DOMAIN 203 210 ZINC-BINDING (BY SIMILARITY).  
 FT DOMAIN 313 345 HEPARIN-BINDING (BY SIMILARITY).  
 FT DOMAIN 413 444 HEPARIN-BINDING (BY SIMILARITY).  
 FT DOMAIN 445 462 COLLAGEN-BINDING (BY SIMILARITY).  
 FT DOMAIN 263 271 POLY-GLU.  
 FT DOMAIN 535 538 POLY-SER.  
 FT DOMAIN 601 606 POLY-LEU.  
 FT SITE 166 166 REQUIRED FOR COPPER(II) REDUCTION (BY SIMILARITY).  
 FT SITE 607 618 BASOLATERAL SORTING SIGNAL (BY SIMILARITY).  
 FT SITE 623 624 CLEAVAGE (BY CASPASE-3) (BY SIMILARITY).  
 FT SITE 641 644 ENDOCYTOSIS SIGNAL.  
 FT SITE 643 646 NPXY MOTIF.  
 FT CARBOHYD 464 464 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 554 554 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT MUTAGEN 641 641 Y->G: REDUCED BINDING OF APBB1.  
 FT CONFLICT 17 17 P -> PP (IN REF. 2).  
 SQ SEQUENCE 653 AA; 72750 MW; 56516DC3EA40E4B0 CRC64;  
 Query Match 82.5%; Score 33; DB 1; Length 653;  
 Best Local Similarity 77.8%; Pred. No. 59;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 LLLPLQILL 9  
 Db 21 LLLPLSLL 29  
 RESULT 11  
 CD8A\_PONPY STANDARD; PRT; 198 AA.  
 AC P30433;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE T-cell surface glycoprotein CD8 alpha chain precursor (T-lymphocyte  
 DE differentiation antigen T8/LEU-2).  
 GN CD8A.  
 OS Pongo pygmaeus (Orangutan).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pongo.  
 OX NCBI\_TaxID=9600;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Isolate Jari.  
 RX MEDLINE=92307742; PubMed=1612644;  
 RA Lawlor D.A.; Parham P.;  
 RT "Structure of CD8 alpha and beta chains of the orangutan: novel  
 RT patterns of mRNA splicing encoding hingeless polypeptides.";  
 RL Immunogenetics 36:121-125(1992).  
 CC -!- FUNCTION: Identifies cytotoxic/suppressor T-cells that interact  
 CC with MHC class I bearing targets. CD8 is thought to play a role in  
 CC the process of T-cell mediated killing. CD8 alpha chains binds to  
 CC class I MHC molecules alpha-3 domains.  
 CC -!- SUBUNIT: In general, heterodimer of an alpha and a beta chain  
 CC linked by two disulfide bonds. Can also form homodimers.  
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -!- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X60223; CAA42784.1; -;

DR PIR; S25656; S25656.  
 DR HSP; P01732; LCD8.  
 DR GO; GO:0042101; C:T-cell receptor complex; ISS.  
 DR GO; GO:0015026; F:coreceptor activity; ISS.  
 DR GO; GO:0042288; F:MHC class I protein binding; ISS.  
 DR GO; GO:0005515; F:protein binding; ISS.  
 DR GO; GO:0006955; P:immune response; ISS.  
 DR GO; GO:0042110; P:T-cell activation; ISS.  
 DR GO; GO:0007169; P:transmembrane receptor protein tyrosine kin. .; ISS.  
 DR InterPro; IPR007110; IG-like.  
 DR Pfam; PF00047; IG\_1.  
 DR SMART; SM00406; IGV; 1.  
 DR PROSITE; PS00835; IG LIKE; 1.  
 KW Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;  
 KW Immune response; Signal.  
 FT SIGNAL 1 21  
 FT CHAIN 22 198 BY SIMILARITY.  
 FT T-CELL SURFACE GLYCOPROTEIN CD8 ALPHA  
 FT CHAIN.  
 FT DOMAIN 22 145 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 146 166 POTENTIAL.  
 FT DOMAIN 167 198 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 22 135 IG-LIKE V-TYPE.  
 FT DISULFID 43 115 BY SIMILARITY.  
 SQ SEQUENCE 198 AA; 22099 MW; F3EC093EAD805561 CRC64;  
 Query Match 80.0%; Score 32; DB 1; Length 198;  
 Best Local Similarity 77.8%; Pred. No. 30;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 LLLPLQILL 9  
 Db 8 LLLPLALL 16  
 RESULT 12  
 CD8A\_HUMAN STANDARD; PRT; 235 AA.  
 ID CD8A\_HUMAN  
 AC P01732;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE T-cell surface glycoprotein CD8 alpha chain precursor (T-lymphocyte  
 DE differentiation antigen T8/Leu-2).  
 GN CD8A OR MAL.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=85099337; PubMed=3871356;  
 RX Littman D.R.; Thomas Y.; Maddon P.J.; Chess L.; Axel R.;  
 RA "The isolation and sequence of the gene encoding T8: a molecule  
 RA defining functional classes of T lymphocytes.";  
 RL Cell 40:237-246(1985).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=86103103; PubMed=3936473;  
 RA Farnes J.R.; Sizer K.C.; Sukhatme V.P.; Hunkapiller T.;  
 RT "Structure of Leu-2/T8 as deduced from the sequence of a cDNA clone.";  
 RL Behring Inst. Mitt. 77:48-55(1985).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=85124510; PubMed=3918796;  
 RA Sukhatme V.P.; Sizer K.C.; Vollmer A.C.; Hunkapiller T.;  
 RA Farnes J.R.;  
 RT "The T cell differentiation antigen Leu-2/T8 is homologous to  
 RT immunoglobulin and T cell receptor variable regions.";  
 RL Cell 40:591-597(1985).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90035142; PubMed=2509342;



J. Biol. Chem. 262:7451-7454(1987).  
 [4]  
 X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF 124-238.  
 MEDLINE=92086855; PubMed=1712141;  
 Weis W.I., Kahn R., Fournier K., Drickamer K., Hendrickson W.A.;  
 "Structure of the calcium-dependent lectin domain from a rat mannose-  
 binding protein determined by MAD phasing";  
 Science 254:1608-1615(1991).  
 [5]  
 X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS) OF 124-238.  
 MEDLINE=93063338; PubMed=1436090;  
 Weis W.I., Drickamer K., Hendrickson W.A.;  
 "Structure of a C-type mannose-binding protein complexed with an  
 oligosaccharide";  
 Nature 360:127-134 (1992).  
 [6]  
 X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF 84-238.  
 MEDLINE=95219384; PubMed=7704532;  
 Weis W.I., Drickamer K.;  
 "Trimeric structure of a C-type mannose-binding protein";  
 Structure 2:1227-1240(1994).  
 [7]  
 X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 90-238.  
 MEDLINE=99119227; PubMed=9922165;  
 Ng K.K.-S., Park-Snyder S., Weis W.I.;  
 "Ca2+-dependent structural changes in C-type mannose-binding  
 proteins";  
 Biochemistry 37:17965-17976(1998).  
 CC -1- FUNCTION: Binds mannose and N-acetylglucosamine in a calcium-  
 dependent manner. Is capable of host defense against pathogens, by  
 activating the classical complement pathway independently of the  
 antibody.  
 CC -1- SUBUNIT: Oligomeric complex of 6 set of homotrimers.  
 CC -1- SUBCELLULAR LOCATION: THIS PROTEIN IS LOCATED INTRACELLULAR, MOST  
 PREDOMINANTLY IN THE ROUGH ENDOPLASMIC RETICULUM AND IN THE GOLGI  
 APPARATUS, AND LITTLE OR NO MBP IS PRESENT IN PLASMA MEMBRANES AND  
 LYSOSOMES. THE MBP IS EXCLUSIVELY LOCALIZED IN THE CISTERNA SPACE  
 OF THE ORGANELLES, PROBABLY AS A LOOSELY BOUND MEMBRANE PROTEIN.  
 CC -1- SIMILARITY: Contains 1 collagenous domain.  
 CC -1- SIMILARITY: Contains 1 C-type lectin family domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 EMBL; M14105; AAA38781.1; -  
 DR EMBL; M14104; AAA38781.1; JOINED.  
 DR PIR; B24791; LNRTWA.  
 DR PDB; 1MGB; 15-JAN-92.  
 DR PDB; 2MSB; 31-OCT-93.  
 DR PDB; 1AFA; 03-APR-96.  
 DR PDB; 1AFB; 03-APR-96.  
 DR PDB; 1AFD; 03-APR-96.  
 DR PDB; 1RTM; 07-FEB-95.  
 DR PDB; 1KMB; 12-FEB-97.  
 DR PDB; 2KMB; 12-FEB-97.  
 DR PDB; 3KMB; 12-FEB-97.  
 DR PDB; 4KMB; 12-FEB-97.  
 DR PDB; 1YMT; 10-JUN-96.  
 DR PDB; 1BCH; 17-JUN-98.  
 DR PDB; 1BCU; 17-JUN-98.  
 DR PDB; 1BUU; 09-SEP-98.  
 DR PDB; 1FIF; 10-JAN-01.  
 DR PDB; 1FTH; 10-JAN-01.  
 DR PDB; 1KWT; 05-JUL-02.  
 DR PDB; 1KWJ; 05-JUL-02.  
 DR PDB; 1KWV; 05-JUL-02.  
 DR PDB; 1KWZ; 05-JUL-02.  
 DR PDB; 1KWY; 05-JUL-02.

DR PDB; 1KWY; 05-JUL-02.  
 DR PDB; 1KWZ; 05-JUL-02.  
 DR PDB; 1KX0; 05-JUL-02.  
 DR PDB; 1KX1; 05-JUL-02.  
 DR InterPro; IPR008160; Collagen.  
 DR InterPro; IPR001304; Lectin C.  
 DR Pfam; PF01391; Collagen; 1.  
 DR Pfam; PF00059; Lectin C; 1.  
 DR SMART; SM00034; CLECT; 1.  
 DR PROSITE; PS00615; C-TYPE LECTIN\_1; 1.  
 DR PROSITE; PS00411; C-TYPE LECTIN\_2; 1.  
 KW Complement pathway; Membrane; Mannose-binding; Calcium; Repeat;  
 KW Signal; Collagen; Lectin; Glycoprotein; Hydroxylation; 3D-structure.  
 FT SIGNAL 1 17  
 FT CHAIN 18 238  
 FT DOMAIN 39 88  
 FT DOMAIN 143 238  
 FT MOD\_RES 43 43  
 FT MOD\_RES 61 61  
 FT MOD\_RES 67 67  
 FT MOD\_RES 73 73  
 FT MOD\_RES 78 78  
 FT DISULFID 145 234  
 FT DISULFID 212 226  
 FT CONFLICT 156 156  
 FT HELIX 91 119  
 FT TURN 120 121  
 FT TURN 124 125  
 FT TURN 138 136  
 FT STRAND 138 147  
 FT HELIX 148 149  
 FT TURN 151 152  
 FT STRAND 151 152  
 FT HELIX 158 168  
 FT STRAND 172 177  
 FT TURN 182 183  
 FT STRAND 186 187  
 FT TURN 188 189  
 FT STRAND 192 192  
 FT STRAND 198 198  
 FT TURN 200 201  
 FT TURN 207 208  
 FT STRAND 212 215  
 FT TURN 217 218  
 FT STRAND 221 224  
 FT TURN 226 227  
 FT STRAND 230 236  
 SQ SEQUENCE 238 AA; 25308 MW; 1A927492B8A8CB3D CRC64;  
 Query Match 80.0%; Score 32; DB 1; Length 238;  
 Best Local Similarity 77.8%; Pred.No.36;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 LLLPLQLL 9  
 Db 2 LLLPLVL 10  
 RESULT 14  
 CD8A\_CANFA  
 ID\_CD8A\_CANFA STANDARD; PRT; 239 AA.  
 AC P33706;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE T-cell surface glycoprotein CD8 alpha chain precursor.  
 GN CD8A.  
 OS Canis familiaris (Dog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
 OX NCBI\_TaxID=9615;  
 RN [1]\_TaxID=9615;  
 RC SEQUENCE FROM N.A.  
 RP STRAIN=Beagle; TISSUE=Thymus;

```

X MEDLINE=94378217; PubMed=8091416;
YA Gorman S.D., Frevin M.R., Cobbold S.P., Waldmann H.;
RT "Isolation and expression of cDNA encoding the canine CD4 and CD8
RT alpha antigens."
RL Tissue Antigens 43:184-188(1994).
CC
CC -!- FUNCTION: Identifies cytotoxic/suppressor T-cells that interact
CC with MHC class I bearing targets. CD8 is thought to play a role in
CC the process of T-cell mediated killing. CD8 alpha chains binds to
CC class I MHC molecules alpha-3 domains.
CC -!- SUBUNIT: In general heterodimer of an alpha and a beta chain
CC linked by two disulfide bonds. Can also form homodimers.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L14287; AAB02294.1;
CC HSP; P01732; ICD8
CC GO; GO:0042101; C: cell receptor complex; ISS.
CC GO; GO:0015026; F: coreceptor activity; ISS.
CC GO; GO:0042288; F: MHC class I protein binding; ISS.
CC GO; GO:0005515; F: protein binding; ISS.
CC GO; GO:0006955; P: immune response; ISS.
CC GO; GO:0042110; P: T-cell activation; ISS.
CC GO; GO:0007169; P: transmembrane receptor protein tyrosine kin. . . ; ISS.
CC InterPro; IPR007110; IG-like.
CC InterPro; IPR003596; IG_LV.
CC Pfam; PF00047; 19; 1.
CC SMART; SM00406; IGV; 1.
CC PROSITE; PS00835; IG_LIKE; 1.
CC Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;
KW Immune response; Signal.
FT SIGNAL 1 21
FT CHAIN 22 239
FT DOMAIN 22 186
FT TRANSMEM 187 210
FT DOMAIN 211 239
FT DOMAIN 25 139
FT DISULFID 46 119
FT CARBOHYD 156 156
FT SEQUENCE 239 AA; 26036 MW; 1018579779A5CB7B CRC64;
SQ
Query Match 80.0%; Score 32; DB 1; Length 239;
Best Local Similarity 77.8%; Pred. No. 36;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

```

RX MEDLINE=91302311; PubMed=2071582;
RA Caughey G.H., Zerweck E.H., Vanderslice P.;
RT "Structure, chromosomal assignment, and deduced amino acid sequence
RT of a human gene for mast cell chymase."
RL J. Biol. Chem. 266:12956-12963(1991).
CC
CC [2]
CC SEQUENCE FROM N.A.
CC TISSUE=Heart;
CC MEDLINE=91373329; PubMed=1894611;
CC Urata H., Kinoshita A., Perez D.M., Misono K.S., Bumpus F.M.,
CC Graham R.M., Hsueh A.;
CC "Cloning of the gene and cDNA for human heart chymase."
CC J. Biol. Chem. 266:17173-17179(1991).
CC [3]
CC SEQUENCE OF 22-247 FROM N.A.
CC MEDLINE=93285916; PubMed=8495723;
CC Sukenaga Y., Kido H., Neki A., Enomoto M., Ishida K., Takagi K.,
CC Katunuma N.;
CC "Purification and molecular cloning of chymase from human tonsils."
CC FEBS Lett. 323:119-122(1993).
CC [4]
CC SEQUENCE OF 26-60 FROM N.A.
CC TISSUE=Placenta;
CC MEDLINE=91264818; PubMed=2049082;
CC Jenne D.E., Tschopp J.;
CC "Angiotensin II-forming heart chymase is a mast-cell-specific
CC enzyme."
CC Biochem. J. 276:567-568(1991).
CC [5]
CC X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
CC MEDLINE=98062898; PubMed=9400368;
CC McGrath W.E., Mirzadegan T., Schmidt B.F.;
CC "Crystal structure of phenylmethanesulfonyl fluoride-treated human
CC chymase at 1.9 A."
CC J. Mol. Biol. 286:163-173(1999).
CC [6]
CC X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).
CC MEDLINE=99134396; PubMed=9931257;
CC Pereira P.J.P., Wang Z.-M., Rubin H., Huber R., Bode W.,
CC Schechter N.M., Strobl S.;
CC "The 2.2-A crystal structure of human chymase in complex with
CC succinyl-Ala-Ala-Pro-Phe-chloromethylketone: structural explanation
CC for its dipeptidyl carboxypeptidase specificity."
CC J. Mol. Biol. 286:163-173(1999).
CC [7]
CC ERRATUM.
CC Pereira P.J.P., Wang Z.-M., Rubin H., Huber R., Bode W.,
CC Schechter N.M., Strobl S.;
CC J. Mol. Biol. 286:817-817(1999).
CC -!- FUNCTION: Major secreted protease of mast cells with suspected
CC roles in vasoactive peptide generation, extracellular matrix
CC degradation, and regulation of gland secretion.
CC -!- CATALYTIC ACTIVITY: Preferential cleavage: Phe-Xaa > Tyr-Xaa >
CC Trp-Xaa > Leu-Xaa.
CC -!- SUBCELLULAR LOCATION: Mast cell granules.
CC -!- TISSUE SPECIFICITY: Mast cells in lung, heart, skin and placenta.
CC -!- SIMILARITY: Belongs to peptidase family S1. Granzyme subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M69137; AAB52021.1;
CC EMBL; M64269; AAB52020.1;
CC EMBL; M59136; AAB52019.1;
CC EMBL; X59072; CAA41796.1;
CC EMBL; S61334; AAB26828.1;
CC PIR; A40967; KYHUCM.
CC PDB; 1KL7; 13-JAN-99.

```

## RESULT 15

MCT1\_HUMAN

ID MCT1\_HUMAN

STANDARD; PRT; 247 AA.

AC F23946; Q16018;

DT 01-MAR-1992 (Rel. 21, Created)

DT 01-MAR-1992 (Rel. 21, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE Chymase precursor (EC 3.4.21.39) (Mast cell protease I).

DE CMA1 OR CVM OR CVH.

OS Homo sapiens (Human)

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

PDB; 1PJP; 02-MAR-99.  
PDB; 1NN6; 08-APR-03.  
MEROPS; S01.140; -.  
Genew; HGNC:2097; CMA1.  
MIM; 118938; -.  
GO; GO:0004252; F:serine-type endopeptidase activity; TAS.  
InterPro; IPR009003; Cys\_Ser\_trypsin.  
InterPro; IPR001254; Peptidase\_S1.  
InterPro; IPR001314; Peptidase\_S1A.  
Pfam; PF00089; trypsin; 1.  
PRINTS; PR00722; CHYMOTRYPSIN.  
SMART; SM00020; TRYP\_SPC; 1.  
PROSITE; PS00240; TRYPSIN\_DOM; 1.  
PROSITE; PS00134; TRYPSIN\_HIS; 1.  
PROSITE; PS00135; TRYPSIN\_SER; 1.  
Hydrolase; Serine protease; Zymogen; Signal;  
3D-structure; Polymorphism.  
FT SIGNAL 19  
FT PROPEP 20 21 ACTIVATION PEPTIDE.  
FT CHAIN 22 247 CHYMASE.  
FT ACT SITE 66 66  
FT ACT SITE 110 110  
FT ACT SITE 203 203  
FT DISULFID 51 67  
FT DISULFID 144 209  
FT DISULFID 175 188  
FT CARBOHYD 80 80  
FT CARBOHYD 103 103  
FT VARIANT 46 46  
FT VARIANT 66 66  
FT CONFLICT 28 28  
FT STRAND 23 23  
FT STRAND 26 27  
FT TURN 30 31  
FT TURN 34 35  
FT STRAND 36 42  
FT STRAND 48 57  
FT TURN 58 59  
FT STRAND 60 63  
FT STRAND 65 67  
FT HELIX 71 76  
FT STRAND 80 80  
FT TURN 81 82  
FT TURN 86 87  
FT STRAND 89 98  
FT TURN 100 101  
FT TURN 104 106  
FT STRAND 112 116  
FT STRAND 123 123  
FT TURN 124 125  
FT STRAND 126 126  
FT TURN 134 135  
FT TURN 140 141  
FT STRAND 143 148  
FT STRAND 151 151  
FT STRAND 158 158  
FT STRAND 161 161  
FT STRAND 163 170  
FT HELIX 172 175  
FT TURN 176 177  
FT TURN 179 180  
FT TURN 183 185  
FT STRAND 186 189  
FT TURN 192 193  
FT STRAND 197 197  
FT TURN 200 201  
FT TURN 203 204  
FT STRAND 206 209  
FT TURN 210 211  
FT STRAND 212 219  
FT TURN 222 223

N-LINKED (GLCNAC. . .) (POTENTIAL).  
N-LINKED (GLCNAC. . .) (POTENTIAL).  
G -> R (in dbSNP:5246).  
/FTid=VAR\_011770.  
H -> R (in dbSNP:5247).  
/FTid=VAR\_011771.  
C -> S (IN REF. 3).

FT STRAND 228 232  
FT HELIX 233 236  
FT HELIX 237 247  
SQ SEQUENCE 247 AA; 27325 MW; DC1454A049ED6B00 CRC64;  
Query Match 80.0%; Score 32; DB 1; Length 247;  
Best Local Similarity 77.8%; Pred. No. 37;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 1 LLLPLQILL 9  
Db 2 LLLPLPLLL 10

Search completed: March 1, 2004, 17:29:56  
Job time : 7 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

AM protein - protein search, using sw model

Run on: March 1, 2004, 17:20:41 ; Search time 30.3333 Seconds  
(without alignments)  
93.615 Million cell updates/sec

Title: US-09-905-083-33

Perfect score: 40

Sequence: 1 LLLPLQLL 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25:  
1: sp\_archaea:  
2: sp\_bacteria:  
3: sp\_fungi:  
4: sp\_human:  
5: sp\_invertebrate:  
6: sp\_mammal:  
7: sp\_mhc:  
8: sp\_organelle:  
9: sp\_phase:  
10: sp\_plant:  
11: sp\_rodent:  
12: sp\_virus:  
13: sp\_vertebrate:  
14: sp\_unclassified:  
15: sp\_rvirus:  
16: sp\_bacteriap:  
17: sp\_archaeap:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	40	100.0	253	Q8N5N9	Q8N5N9 homo sapien
2	38	95.0	73	Q8VCA9	Q8VCA9 mus musculus
3	38	95.0	104	Q8R5D6	Q8R5D6 mus musculus
4	37	92.5	208	Q21527	Q21527 clethrionom
5	36	90.0	146	Q9V2D5	Q9V2D5 pyrococcus
6	34	85.0	162	Q91T40	Q91T40 lumpy skin
7	34	85.0	162	Q91M24	Q91M24 lumpy skin
8	34	85.0	277	Q9HP13	Q9HP13 halobacteri
9	34	85.0	369	Q863A1	Q863A1 bos taurus
10	34	85.0	754	Q8ZQC3	Q8ZQC3 salmonella
11	34	85.0	754	Q8Z802	Q8Z802 salmonella
12	34	85.0	845	Q97H76	Q97H76 clostridium
13	33	82.5	109	Q8BJK7	Q8BJK7 mus musculus
14	33	82.5	196	Q21527	Q21527 clethrionom
15	33	82.5	196	Q9D226	Q9D226 mus musculus
16	33	82.5	282	Q7T268	Q7T268 mycobacteri

17	33	82.5	294	16	Q53979	Q53979 mycobacteri
18	33	82.5	370	16	Q8VHY9	Q8VHY9 bruceella me
19	33	82.5	370	16	Q8FX1	Q8FX1 bruceella su
20	33	82.5	399	16	Q83K08	Q83K08 shigella fl
21	33	82.5	413	16	Q8XFY3	Q8XFY3 salmonella
22	33	82.5	426	16	Q9HW63	Q9HW63 pseudomonas
23	33	82.5	428	16	Q8X9C3	Q8X9C3 escherichia
24	33	82.5	470	8	Q8SEA5	Q8SEA5 guillardia
25	33	82.5	475	17	Q26479	Q26479 methanobact
26	33	82.5	516	16	Q82H48	Q82H48 streptomyce
27	33	82.5	547	16	Q8E1Y2	Q8E1Y2 shewanella
28	32	80.0	58	17	Q8U289	Q8U289 pyrococcus
29	32	80.0	72	12	Q8VBB6	Q8VBB6 white spot
30	32	80.0	79	12	Q8VB59	Q8VB59 white spot
31	32	80.0	102	11	Q8BRV0	Q8BRV0 mus musculu
32	32	80.0	116	16	Q8FQ09	Q8FQ09 leptospira
33	32	80.0	150	16	Q9CKE2	Q9CKE2 pasteurella
34	32	80.0	182	5	Q8MZ16	Q8MZ16 mamestra br
35	32	80.0	182	5	Q8MZK7	Q8MZK7 plodia inte
36	32	80.0	182	5	Q8MZV6	Q8MZV6 spodoptera
37	32	80.0	182	5	Q8MZZ6	Q8MZZ6 cydia pomon
38	32	80.0	198	4	Q13970	Q13970 homo sapien
39	32	80.0	235	4	Q8TAW8	Q8TAW8 homo sapien
40	32	80.0	235	4	Q96QR6	Q96QR6 homo sapien
41	32	80.0	235	6	Q9XSM6	Q9XSM6 salmire sci
42	32	80.0	242	16	Q7VHT6	Q7VHT6 helicobacte
43	32	80.0	249	16	Q8X6D3	Q8X6D3 escherichia
44	32	80.0	256	11	Q92ZG5	Q92ZG5 mus musculu
45	32	80.0	264	16	Q8UIE2	Q8UIE2 agrobacteri

## ALIGNMENTS

### RESULT 1

Q8N5N9	PRELIMINARY;	PRT;	253 AA.
ID	Q8N5N9		
AC	Q8N5N9;		
DT	01-OCT-2002 (TREMELrel. 22, Created)		
DT	01-OCT-2002 (TREMELrel. 22, Last sequence update)		
DT	01-OCT-2003 (TREMELrel. 25, Last annotation update)		
DE	Kalikrein 7 (chymotryptic, stratum corneum).		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Skin.		
RA	Strausberg R.		
RL	Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.		
CC	-(- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.		
DR	EMBL, BC032005; AH32005.1; -		
DR	GO; GO:0004263; F:chymotrypsin activity; IEA.		
DR	GO; GO:0008233; F:peptidase activity; IEA.		
DR	GO; GO:0004295; F:trypsin activity; IEA.		
DR	GO; GO:0006508; P:proteolysis and peptidolysis; IEA.		
DR	InterPro; IPR009003; Cys Ser trypsin.		
DR	InterPro; IPR001254; peptidase S1.		
DR	InterPro; IPR001314; peptidase_S1A.		
DR	Pfam; PF00089; trypsin; 1		
DR	PRINTS; PR00722; CHYMOTRYPSIN.		
DR	SMART; SM00020; Tryp_SPC; 1.		
DR	PROSITE; PS00240; TRYPsin_DOM; 1.		
DR	PROSITE; PS00134; TRYPsin_HIS; 1.		
DR	PROSITE; PS00135; TRYPsin_SER; 1.		
KW	Hydrolase; Protease; Serine protease.		
SK	SEQUENCE 253 AA; 27608 MW; 2D68B6A41B22A668 CRC64;		

Query Match 100.0%; Score 40; DB 4; Length 253;  
Best Local Similarity 100.0%; Pred. No. 5-2;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 LLLPLQLLL 9  
b 5 LLLPLQLLL 13

RESULT 2  
8VCAG PRELIMINARY; PRT; 73 AA.  
C Q8VCA9  
T 01-MAR-2002 (TrEMBLrel. 20, Created)  
T 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
T 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
DE Similar to RIKEN cDNA 1110002023 gene.  
S Mus musculus (Mouse).  
X Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
X Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
X NCBI\_TaxID=10090;  
X [1]  
R SEQUENCE FROM N.A.  
R TISSUE=Salivary Gland;  
R Strausberg R.;  
R Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.  
R EMBL; BC021345; AAB21345.1; -  
S Q SEQUENCE 73 AA; 7819 MW; 938E533998F3C11 CRC64;

Query Match 95.0%; Score 38; DB 11; Length 73;  
Best Local Similarity 88.9%; Pred. No. 4, 2;  
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLLPLQLLL 9  
b 6 LLLPLQLLL 14

RESULT 3  
Q8R5D6 PRELIMINARY; PRT; 104 AA.  
AC Q8R5D6  
T 01-JUN-2002 (TrEMBLrel. 21, Created)  
T 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
T 01-JUN-2002 (TrEMBLrel. 24, Last annotation update)  
DE Similar to RIKEN cDNA 1110002023 gene.  
GN FKBP11 OR 1110002023RIK.  
OS Mus musculus (Mouse).  
X Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
X Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
X NCBI\_TaxID=10090;  
X [1]  
R SEQUENCE FROM N.A.  
R Strausberg R.;  
R Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
R EMBL; BC022300; AAB22300.1; -  
R MGD; MG11913370; FKBP11.  
R GO; GO:0006457; P:protein folding; IEA.  
R InterPro; IPR001179; FKBP\_PPase.  
R Pfam; PF00254; FKBP; 1.  
R PROSITE; PS00453; FKBP\_PPase\_1; 1.  
R PROSITE; PS00059; FKBP\_PPase\_3; 1.  
S Q SEQUENCE 104 AA; 11085 MW; 0534D57467566914 CRC64;

Query Match 95.0%; Score 38; DB 11; Length 104;  
Best Local Similarity 88.9%; Pred. No. 5, 8;  
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLLPLQLLL 9  
b 6 LLLPLQLLL 14

RESULT 4  
Q21527 PRELIMINARY; PRT; 208 AA.  
AC Q21527;  
AC Q21527;

DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE NADH dehydrogenase subunit 4 (EC 1.6.5.3) (NADH-ubiquinone  
oxidoreductase chain 4) (Fragment).  
GN ND4  
OS Clethrionomys gapperi (Southern red-backed vole).  
OC Mitochondrion.  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Arvicolinae;  
OC Clethrionomys.  
OX NCBI\_TaxID=56223;  
X [1]  
R SEQUENCE FROM N.A.  
R MEDLINE=98152303; PubMed=9491603;  
R Engel S.R.; Hogan K.M.; Taylor J.F.; Davis S.K.;  
R "Molecular systematics and paleobiogeography of the South American  
R sigmodontine rodents";  
R Mol. Biol. Evol. 15:35-49(1998).  
CC -/- CATALYTIC ACTIVITY: NADH + UBIQUINONE = NAD(+) + UBIQUINOL.  
DR EMBL; U83808; AAB87168.1; -  
DR GO; GO:0005739; C:mitochondrion; IEA.  
DR GO; GO:0008137; F:NADH dehydrogenase (ubiquinone) activity; IEA.  
DR GO; GO:0016491; F:oxidoreductase activity; IEA.  
DR GO; GO:0006120; P:mitochondrial electron transport, NADH to u...; IEA.  
DR InterPro; IPR003918; NADHub\_oxred4.  
DR InterPro; IPR001750; Oxidored\_g1.  
DR InterPro; IPR000260; Oxidored\_g5\_N.  
DR Pfam; PF00361; oxidored\_g1; 1.  
DR Pfam; PF01059; oxidored\_g5\_N; 1.  
DR PRINTS; PR01437; NUOXDRDTASE4.  
DR NAD; Oxidoreductase; Ubiquinone; Mitochondrion.  
FT NON TER 208  
FT SEQUENCE 208 AA; 23967 MW; 8AF1788697AED6A2 CRC64;

Query Match 92.5%; Score 37; DB 8; Length 208;  
Best Local Similarity 88.9%; Pred. No. 17;  
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLLPLQLLL 9  
b 98 LLLPLQLLL 106

RESULT 5  
Q9V2D5 PRELIMINARY; PRT; 146 AA.  
AC Q9V2D5  
T 01-MAY-2000 (TrEMBLrel. 13, Created)  
T 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
T 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Hypothetical protein PYRAB01390.  
GN PYRAB01390 OR PAB0088.  
OS Pyrococcus abyssi.  
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;  
OX Pyrococcus.  
OX NCBI\_TaxID=29292;  
X [1]  
R SEQUENCE FROM N.A.  
R STRAIN=GES / Orsay;  
R Hailig R.;  
R "Pyrococcus abyssi genome sequence: insights into archaeal chromosome  
R structure and evolution";  
R Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
R EMBL; AJ248283; CAB49063.1; -  
R PIR; H75201; H75201.  
DR InterPro; IPR008537; DUF819.  
DR Pfam; PF05684; DUF819; 1.  
DR Hypothetical protein; Complete proteome.  
S Q SEQUENCE 146 AA; 16092 MW; 7182941371258C1F CRC64;

Query Match 90.0%; Score 36; DB 17; Length 146;  
Best Local Similarity 77.8%; Pred. No. 20;

Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

1 LLLPLQIILL 9  
 67 LLLPLQIILL 75

RESULT 6  
 Q91M24 PRELIMINARY; PRT; 162 AA.  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 E Hypothetical protein (LAP/BHD-finger protein).  
 GN L01010.  
 OS Lumpy skin disease virus (LSDV).  
 NC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;  
 NC Capripoxvirus.  
 NC NCBI\_TaxID=59509;  
 UN [1]

SEQUENCE FROM N.A.  
 RC STRAIN=Neethling;  
 RA Stipinovich C., Vreede F.T., Taljaard L.C.F., Zsak A., Viljoen G.J., Rock D.L.;  
 RA Viljoen G.J.;  
 RT "Molecular characterization of important regions of the Lumpy skin  
 RT disease virus genome";  
 RT Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
 RN [2]

SEQUENCE FROM N.A.  
 RC STRAIN=Neethling vaccine LW 1959;  
 RA Kara P.D., Afonso C.L., Wallace D.B., Kutish G.F., Stipinovich C.,  
 RA Lu Z., Vreede F.T., Taljaard L.C.F., Zsak A., Viljoen G.J., Rock D.L.;  
 RT "Molecular characterization of the South African vaccine strain and  
 RT the field isolate of lumpy skin disease virus";  
 RT Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.  
 DR EMBL; AF336128; AAK43550.1; -;  
 DR EMBL; AF409138; AAN02734.1; -;  
 DR InterPro: IPR001841; Znf\_ring.  
 DR Pfam: PF00097; zf-CHC4\_1.  
 KW Hypothetical protein; Metal-binding; Zinc; Zinc-finger.  
 SQ SEQUENCE 162 AA; 18782 MW; 59914A080F729EE CRC64;

Query Match 85.0%; Score 34; DB 12; Length 162;  
 Best Local Similarity 88.9%; Pred. No. 55;  
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQIILL 9  
 90 LLLPLQIILL 98

RESULT 7  
 Q91M24 PRELIMINARY; PRT; 162 AA.  
 AC Q91M24;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE LSDV010 LAP/BHD-finger protein.  
 GN LSDV010 OR L01010.  
 OS Lumpy skin disease virus (LSDV).  
 NC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;  
 NC Capripoxvirus.  
 NC NCBI\_TaxID=59509;  
 UN [1]

SEQUENCE FROM N.A.  
 RC STRAIN=Neethling 2490;  
 RC MEDLINE=21329495; PubMed=11435593;  
 RX Tulman E.R., Afonso C.L., Lu Z., Zsak A., Kutish G.F., Rock D.L.;  
 RT "Genome of lumpy skin disease virus";  
 RT J. Virol. 75:7122-7130(2001).

SEQUENCE FROM N.A.  
 RC STRAIN=Neethling 2490;  
 RA Tulman E.R., Afonso C.L., Lu Z., Zsak A., Kutish G.F., Rock D.L.;  
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
 RN [3]

SEQUENCE FROM N.A.  
 RC STRAIN=Neethling Warmbaths LW;  
 RA Kara P.D., Afonso C.L., Wallace D.B., Kutish G.F., Stipinovich C.,  
 RA Lu Z., Vreede F.T., Taljaard L.C.F., Zsak A., Viljoen G.J., Rock D.L.;  
 RT "Molecular characterization of the South African vaccine strain and  
 RT the field isolate of lumpy skin disease virus";  
 RT Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.  
 DR EMBL; AF325528; AAK4971.1; -;  
 DR EMBL; AF409137; AAN02576.1; -;  
 DR InterPro: IPR001841; Znf\_ring.  
 DR Pfam: PF00097; zf-CHC4\_1.  
 KW Metal-binding; Zinc; Zinc-finger.  
 SQ SEQUENCE 162 AA; 18835 MW; 0EA24745C3818222 CRC64;

Query Match 85.0%; Score 34; DB 12; Length 162;  
 Best Local Similarity 88.9%; Pred. No. 55;  
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQIILL 9  
 90 LLLPLQIILL 98

RESULT 8  
 Q9HP13 PRELIMINARY; PRT; 277 AA.  
 AC Q9HP13;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Cytochrome a3 controlling protein.  
 GN CYP OR VNGH623G.  
 OS Halobacterium sp. (strain NRC-1 / ATCC 700922 / JCM 11081).  
 NC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;  
 NC Halobacteriaceae; Halobacterium.  
 NC NCBI\_TaxID=64091;  
 UN [1]

SEQUENCE FROM N.A.  
 RC MEDLINE=20504483; PubMed=11016950;  
 RA Ng W.V., Kennedy S.P., Mahaitas G.G., Berquist B., Pan M.,  
 RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,  
 RA Swartzell S., Weir D., Hall J., Dahl T.A., Melti R., Goo Y.A.,  
 RA Leithausen B., Keller K., Cruz R., Dawson M.J., Hough D.W.,  
 RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,  
 RA Isenbarger T.A., Peck R.F., Fohlschroder M., Spudich J.L., Jung K.-H.,  
 RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,  
 RA Ebhardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.;  
 RT "Genome sequence of Halobacterium species NRC-1";  
 RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).  
 DR EMBL; AEO05072; AAG19884.1; -;  
 DR PTR; H84314; H84314.  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR GO; GO:0006461; P:protein complex assembly; IEA.  
 DR InterPro: IPR003780; COX15\_Ctaa.  
 DR Pfam: PF02628; COX15\_Ctaa; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 277 AA; 29053 MW; 805312376E274B1 CRC64;

Query Match 85.0%; Score 34; DB 17; Length 277;  
 Best Local Similarity 66.7%; Pred. No. 90;  
 Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLLPLQIILL 9  
 97 LLLPLQIILL 105



Query Match 85.0%; Score 34; DB 16; Length 754;

```
Best Local Similarity 77.8%; Pred. No. 2.3e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

1 LLLPLQILL 9
371 LLLPLOVAL 379

RESULT 12
37H76 PRELIMINARY; PRT; 845 AA.
C Q97H76;
I 01-OCT-2001 (TREMBlrel. 18, Created)
I 01-OCT-2001 (TREMBlrel. 18, Last sequence update)
I 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
E Cation transport P-type ATPase.
N CAC2137.
S Clostridium acetobutylicum.
C Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
C Clostridium.
X NCBI_TaxID=1488;
[1]
SEQUENCE FROM N.A.
C STRAIN=ATCC 824 / DSM 792 / VPM B-1787;
C MEDLINE=21359325; PubMed=11466286;
A Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
A Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
A Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
A Bennett G.N., Koonin E.V., Smith D.R.,
A "Genome sequence and comparative analysis of the solvent-producing
RT bacterium Clostridium acetobutylicum."
RL J. Bacteriol. 183:4823-4838 (2001).
DR EMBL; A800714; AAK80095.1; -.
DR PIR; D97163; D97163.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0015662; F:ATPase activity, coupled to transmembrane m...; IEA.
DR GO; GO:0015787; F:hydrolase activity; IEA.
DR GO; GO:0006812; P:cation transport; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR GO; GO:0015992; P:proton transport; IEA.
DR InterPro; IPR001757; ATPase_E1-E2.
DR InterPro; IPR006068; Cation ATPase C.
DR InterPro; IPR004014; Cation ATPase N.
DR InterPro; IPR008250; E1-E2 ATPase_N.
DR InterPro; IPR005834; Hydrolase.
DR InterPro; IPR000695; H ATPase.
DR Pfam; PF00689; Cation ATPase C; 1.
DR Pfam; PF00690; Cation ATPase N; 1.
DR Pfam; PF00122; E1-E2 ATPase; 1.
DR Pfam; PF00702; Hydrolase; 1.
DR PRINTS; PR00139; CATAPASE.
DR PRINTS; PR00120; HATPASE.
DR TIGRFAMs; TIGR01494; ATPase P-type; 5.
DR PROSITE; PS00154; ATPASE_E1_E2; 1.
KW Complete proteome.
SQ SEQUENCE 845 AA; 93779 MW; 401293AA1FF9D757 CRC64;

Query Match 85.0%; Score 34; DB 16; Length 845;
Best Local Similarity 87.5%; Pred. No. 2.5e+02;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 LLLPLQILL 9
DB 670 LLLPLQILL 677

RESULT 13
Q8BJK7 PRELIMINARY; PRT; 109 AA.
AC Q8BJK7
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)

DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Glutamine.
DE Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
OX [1]
RN SEQUENCE FROM N.A.
RP STRAIN=C57BL/6J; TISSUE=Body;
RC MEDLINE=22354683; PubMed=12466851;
RX The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573 (2002).
DR EMBL; AK083529; BAC38942.1; -.
SQ SEQUENCE 109 AA; 12276 MW; 852E75D4571F8DC6 CRC64;

Query Match 92.5%; Score 33; DB 11; Length 109;
Best Local Similarity 86.7%; Pred. No. 60;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9
DB 37 LLLPLQILL 45

RESULT 14
Q9X620 PRELIMINARY; PRT; 196 AA.
AC Q9X620
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE CORF.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
OX [1]
RN SEQUENCE FROM N.A.
RP STRAIN=LT2;
RC Smith R.L., Ahuga D., Thacker L.K., Maguire M.E.;
RT "Magnesium transport in Salmonella typhimurium: Sequence and
RT characterization of the corB, corC, and corD genes."
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF130857; AAD31438.1; -.
DR InterPro; IPR002550; CBS.
DR Pfam; PF01595; DUF21; 1.
SQ SEQUENCE 196 AA; 21859 MW; B06659F7E5BAA17A CRC64;

Query Match 92.5%; Score 33; DB 2; Length 196;
Best Local Similarity 77.8%; Pred. No. 1e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9
DB 123 LLLPLQILL 131

RESULT 15
Q9D226 PRELIMINARY; PRT; 196 AA.
AC Q9D226
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE A030007E19Rik protein.
GN A030007E19Rik.
GN Mus musculus (Mouse).
OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
```

```

( NCBI_TaxID=10090;
[1]
? SEQUENCE FROM N.A.
? STRAIN=CS7BL/6J; TISSUE=Skin;
? MEDLINE=21085660; PubMed=11217851;
A Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
A Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
A Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
A Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,
A Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
A Fleischmann W., Gaasterland T., Glass C., King B., Kochiwa H.,
A Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pescie G., Quackenbush J.,
A Schirani L.M., Staubli P., Suzuki R., Tomita M., Wagner L., Washio T.,
A Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
A Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.P.,
A Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
A Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
A Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
A Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
A Sasaki H., Sato K., Schoenbach C., Seva T., Shibata Y., Storch K.-F.,
A Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
A Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
A Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK020699; BAB32180.1; -.
DR HSSP; P10969; IWGT.
DR MGD; MGI:1924465; A030007E19Rik.
DR GO; GO:0003754; F:chaperone activity; IEA.
DR GO; GO:0005489; F:electron transporter activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR001450; 4Peas_ferredoxin.
DR PROSITE; PS00198; 4FE4S_FERREDOXIN; 4.
KW Chaperone; Repeat.
SQ
SEQUENCE 196 AA; 17777 MW; 2136738FD496D815 CRC64;

```

```

Query Match      82.5%; Score 33; DB 11; Length 196;
Best Local Similarity 77.8%; Pred. No. 1e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 1 LLLPLQLLL 9
DB 156 LLLPLQLLL 164

```

```

Search completed: March 1, 2004, 17:34:44
Job time : 32.333 secs

```

GanCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

#### 4 protein - protein search, using sw model

an on: March 1, 2004, 17:21:46 ; Search time 11.8889 Seconds  
(without alignments)  
39.081 Million cell updates/sec

title: US-09-905-083-34

erfect score: 47

sequence: 1 VLVNERWVL 9

coring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

searched: 389414 seqs, 51625971 residues

total number of hits satisfying chosen parameters: 389414

minimum DB seq length: 0

Maximum Match 100%

Maximum Match 0%

Listing first 45 summaries

#### Database :

Issued Patents AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A COMB.pcp:\*  
2: /cgn2\_6/ptodata/2/iaa/5B COMB.pcp:\*  
3: /cgn2\_6/ptodata/2/iaa/6A COMB.pcp:\*  
4: /cgn2\_6/ptodata/2/iaa/6B COMB.pcp:\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS COMB.pcp:\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pcp:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	47	100.0	9	3	US-09-502-600-34
2	47	100.0	9	4	US-09-918-243-34
3	47	100.0	154	3	US-09-261-416-7
4	47	100.0	224	3	US-08-944-483-33
5	47	100.0	225	2	US-08-557-146-12
6	47	100.0	225	2	US-09-027-337-4
7	47	100.0	225	2	US-09-154-344-12
8	47	100.0	225	4	US-08-644-600-4
9	47	100.0	225	4	US-08-654-600A-4
10	47	100.0	253	2	US-08-557-146-2
11	47	100.0	253	2	US-08-824-874-3
12	47	100.0	253	2	US-09-154-344-2
13	47	100.0	253	3	US-09-930-188-2
14	47	100.0	253	3	US-09-210-084-3
15	47	100.0	253	4	US-09-764-762-3
16	47	100.0	253	5	PCT-US96-04294-2
17	43	91.5	9	3	US-09-502-600-40
18	43	91.5	9	3	US-09-502-600-50
19	43	91.5	9	4	US-09-918-243-40
20	43	91.5	9	4	US-09-918-243-50
21	39	83.0	228	3	US-08-944-483-44
22	39	83.0	253	6	5223425-8
23	38	80.9	64	1	US-08-485-455D-75
24	38	80.9	64	2	US-08-482-130C-75
25	38	80.9	64	2	US-08-484-211C-75
26	38	80.9	64	3	US-08-906-769-75
27	38	80.9	64	3	US-08-906-616-75

```

28 38 80.9 64 3 US-08-817-795-75 Sequence 75, Appl
29 38 80.9 64 3 US-08-485-443B-75 Sequence 75, Appl
30 38 80.9 64 3 US-08-639-075A-75 Sequence 75, Appl
31 38 80.9 64 3 US-09-012-431-75 Sequence 75, Appl
32 38 80.9 64 3 US-09-012-692-75 Sequence 75, Appl
33 38 80.9 64 3 PCT-US95-14442A-75 Sequence 75, Appl
34 38 80.9 223 1 US-08-485-455D-17 Sequence 17, Appl
35 38 80.9 223 2 US-08-482-130C-17 Sequence 17, Appl
36 38 80.9 223 3 US-08-484-211C-17 Sequence 17, Appl
37 38 80.9 223 3 US-08-817-795-17 Sequence 17, Appl
38 38 80.9 223 3 PCT-US95-14442A-17 Sequence 17, Appl
39 38 80.9 223 5 US-08-485-443B-17 Sequence 17, Appl
40 38 80.9 224 3 US-08-306-769-17 Sequence 17, Appl
41 38 80.9 224 3 US-08-306-616-17 Sequence 17, Appl
42 38 80.9 224 3 US-08-639-075A-17 Sequence 17, Appl
43 38 80.9 224 3 US-09-012-431-17 Sequence 17, Appl
44 38 80.9 224 3 US-09-012-692-17 Sequence 17, Appl
45

```

#### ALIGNMENTS

##### RESULT 1

US-09-502-600-34

; Sequence 34, Application US/09502600A

; Patent No. 6294344

; GENERAL INFORMATION:

; APPLICANT: O'Brien, Timothy J.

; TITLE OF INVENTION: Compositions and Methods for the Early Diagnosis of

; FILE REFERENCE: D6223CIP-C

; CURRENT FILING DATE: 2000-02-11

; PRIOR APPLICATION NUMBER: US/09/502,600A

; PRIOR FILING DATE: 09/039,211

; NUMBER OF SEQ ID NOS: 136

; SEQ ID NO 34

; LENGTH: 9

; TYPE: PRT

; ORGANISM: Homo sapiens

; FEATURE:

; OTHER INFORMATION: Residues 58-66 of the SCCE protein

US-09-502-600-34

Query Match

Best Local Similarity 100.0%; Score 47; DB 3; Length 9;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLVNERWVL 9

Db 1 VLVNERWVL 9

##### RESULT 2

US-09-918-243-34

; Sequence 34, Application US/09918243

; Patent No. 6627403

; GENERAL INFORMATION:

; APPLICANT: O'Brien, Timothy J.

; APPLICANT: Cannon, Martin J.

; APPLICANT: Santin, Alessandro

; TITLE OF INVENTION: Methods for the early diagnosis of ovarian cancer

; FILE REFERENCE: D6223CIP/C/D/CIP

; CURRENT APPLICATION NUMBER: US/09/918,243

; CURRENT FILING DATE: 2001-07-30

; PRIOR APPLICATION NUMBER: US

; PRIOR FILING DATE: 2001-07-13

; NUMBER OF SEQ ID NOS: 136

; SEQ ID NO 34

; LENGTH: 9

; TYPE: PRT

; ORGANISM: Homo sapiens

FEATURE:  
NAME/KEY: CHAIN  
OTHER INFORMATION: Residues 58-66 of the SCCE protein  
3-09-918-243-34

Query Match 100.0%; Score 47; DB 4; Length 9;  
Best Local Similarity 100.0%; Pred. No. 3e+05;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 VLVNERWVL 9  
b 1 VLVNERWVL 9

RESULT 3  
US-09-261-416-7  
Sequence 7, Application US/09261416A  
Patent No. 6291663  
GENERAL INFORMATION:  
APPLICANT: O'Brien, Timothy J.  
APPLICANT: Underwood, Lowell J.  
TITLE OF INVENTION: TAGD-12: A No. 6291663el Transmembrane Serine Protease  
TITLE OF INVENTION: TAGD-12: A No. 6291663el Transmembrane Serine Protease  
FILE REFERENCE: D6192  
CURRENT APPLICATION NUMBER: US/09/261.416A  
CURRENT FILING DATE: 1999-03-03  
NUMBER OF SEQ ID NOS: 14  
SEQ ID NO 7  
LENGTH: 154  
TYPE: PRT  
ORGANISM: Unknown

FEATURE:  
OTHER INFORMATION: Serine protease catalytic domain of stratum corneum  
OTHER INFORMATION: chymotryptic enzyme (Scce) homologous to similar  
OTHER INFORMATION: domain in TAGD-12

US-09-261-416-7

Query Match 100.0%; Score 47; DB 3; Length 154;  
Best Local Similarity 100.0%; Pred. No. 0.29;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLVNERWVL 9  
Db 4 VLVNERWVL 12

RESULT 4  
US-08-944-483-33  
Sequence 33, Application US/08944483  
Patent No. 6232456  
GENERAL INFORMATION:  
APPLICANT: COHEN, MAURICE  
APPLICANT: COLPITTS, TRACEY L.  
APPLICANT: FRIEDMAN, PAULA N.  
APPLICANT: GRANADOS, EDWARD N.  
APPLICANT: KLASS, MICHAEL R.  
APPLICANT: RUSSELL, JOHN C.  
APPLICANT: STEWART, KENT D.  
APPLICANT: STROUPE, STEVEN D.  
TITLE OF INVENTION: NOVEL SERINE PROTEASE REAGENTS  
TITLE OF INVENTION: AND METHODS USEFUL FOR DETECTING AND TREATING DISEASES  
NUMBER OF SEQUENCES: 76  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Abbott Laboratories  
STREET: 100 Abbott Park Road  
CITY: Abbott Park  
STATE: IL  
COUNTRY: USA  
ZIP: 60064-3500  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS  
SOFTWARE: FastSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/944,483  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Becker, Cheryl L.  
REGISTRATION NUMBER: 35,441  
REFERENCE/DOCKET NUMBER: 6183.US.01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 847/935-1729  
TELEFAX: 847/938-2623  
TELEX:  
INFORMATION FOR SEQ ID NO: 33:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 224 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: No. 6232456e  
US-08-944-483-33

Query Match 100.0%; Score 47; DB 3; Length 224;  
Best Local Similarity 100.0%; Pred. No. 0.43;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLVNERWVL 9  
Db 29 VLVNERWVL 37

RESULT 5  
US-08-557-146-12  
Sequence 12, Application US/08557146  
Patent No. 5834290  
GENERAL INFORMATION:  
APPLICANT: Egelrud, Torbjorn  
APPLICANT: Hansson, Lennart  
TITLE OF INVENTION: Recombinant Stratum Corneum Chymotryptic  
TITLE OF INVENTION: Enzyme (SCCE)  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: White & Case, Patent Department  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2787  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/557,146  
FILING DATE: 14-DEC-1995  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Steiner, Richard J.  
REGISTRATION NUMBER: 35,372  
REFERENCE/DOCKET NUMBER: 1103326-181  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 819-8783  
TELEFAX: (212) 354-8113  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 225 amino acids  
TYPE: amino acids  
STRANDEDNESS: single

TOPOLOGY: linear  
MOLECULE TYPE: polypeptide  
-08-557-146-12

Query Match 100.0%; Score 47; DB 2; Length 225;  
Best Local Similarity 100.0%; Pred. No. 0.43;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 VLVNERWVL 9  
|||||||  
30 VLVNERWVL 38

## RESULT 6

3-09-027-337-4  
Sequence 4, Application US/09027337B  
Patent No. 5972616

## GENERAL INFORMATION:

APPLICANT: O'Brien, Timothy J.  
TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease Overexpressed in Breast and Ovarian Carcinomas  
FILE REFERENCE: D6064  
CURRENT APPLICATION NUMBER: US/09/027,337B  
CURRENT FILING DATE: 1998-02-20  
NUMBER OF SEQ ID NOS: 13  
SEQ ID NO 4  
LENGTH: 225  
TYPE: PRT  
ORGANISM: Unknown  
FEATURE:  
OTHER INFORMATION: Serine protease catalytic domain of Scce homologous to similar domain in TADG-15

Query Match 100.0%; Score 47; DB 2; Length 225;  
Best Local Similarity 100.0%; Pred. No. 0.43;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 VLVNERWVL 9  
|||||||  
30 VLVNERWVL 38

## RESULT 7

US-09-154-344-12  
Sequence 12, Application US/09154344  
Patent No. 5981256

## GENERAL INFORMATION:

APPLICANT: Egelrud, Torbjorn  
TITLE OF INVENTION: Recombinant Stratum Corneum Chymotryptic Enzyme (SCCE)  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: White & Case, Patent Department  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2787

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/154,344  
FILING DATE: 16-SEP-1998  
CLASSIFICATION:  
PRIORITY DATA:  
APPLICATION NUMBER: US 08/557,146  
FILING DATE: 14-DEC-1995

## CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:  
NAME: Steiner, Richard J.  
REGISTRATION NUMBER: 35,372  
REFERENCE/DOCKET NUMBER: 1103326-181  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 819-8783  
TELEFAX: (212) 354-8113  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 225 amino acids  
TYPE: amino acids  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: polypeptide  
US-09-154-344-12

Query Match 100.0%; Score 47; DB 2; Length 225;  
Best Local Similarity 100.0%; Pred. No. 0.43;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 VLVNERWVL 9  
|||||||  
30 VLVNERWVL 38

## RESULT 8

US-09-644-600-4  
Sequence 4, Application US/09644600  
Patent No. 6451500

## GENERAL INFORMATION:

APPLICANT: O'Brien, Timothy J.  
TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease Overexpressed in Carcinomas  
FILE REFERENCE: D6064CIP/D  
CURRENT APPLICATION NUMBER: US/09/644,600  
CURRENT FILING DATE: 2000-08-23  
PRIOR APPLICATION NUMBER: 09/421,213  
PRIOR FILING DATE: 1999-10-20  
PRIOR APPLICATION NUMBER: 09/027,337  
PRIOR FILING DATE: 1998-02-20  
NUMBER OF SEQ ID NOS: 98  
SEQ ID NO 4  
LENGTH: 225  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: SCCE  
US-09-644-600-4

Query Match 100.0%; Score 47; DB 4; Length 225;  
Best Local Similarity 100.0%; Pred. No. 0.43;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 VLVNERWVL 9  
|||||||  
30 VLVNERWVL 38

## RESULT 9

US-09-654-600A-4  
Sequence 4, Application US/09654600A  
Patent No. 6649741

## GENERAL INFORMATION:

APPLICANT: O'Brien, Timothy J.  
TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease Overexpressed in Carcinomas  
FILE REFERENCE: D6064CIP/D  
CURRENT APPLICATION NUMBER: US/09/654,600A  
CURRENT FILING DATE: 2000-09-01  
PRIOR APPLICATION NUMBER: 09/421,213

09/027,337  
PRIOR FILING DATE: 1999-10-20  
NUMBER OF SEQ ID NOS: 98  
SEQ ID NO. 4  
LENGTH: 225  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: SCCE  
:-09-554-600A-4

Query Match 100.0%; Score 47; DB 4; Length 225;  
Best Local Similarity 100.0%; Pred. No. 0.43;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 VLVNERWVL 9  
|||||||  
b 30 VLVNERWVL 38

## RESULT 10

US-08-557-146-2  
Sequence 2, Application US/08557146  
Patent No. 5834290

GENERAL INFORMATION:  
APPLICANT: Egelrud, Torbjorn  
TITLE OF INVENTION: Recombinant Stratum Corneum Chymotryptic  
TITLE OF INVENTION: Enzyme (SCCE)  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: White & Case, Patent Department  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2787

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/557,146  
FILING DATE: 14-DEC-1995  
CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:  
NAME: Sterner, Richard J.

REGISTRATION NUMBER: 35,372  
REFERENCE/DOCKET NUMBER: 1103326-181

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 819-8783  
TELEFAX: (212) 354-8113

INFORMATION FOR SEQ ID NO. 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 253 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein

US-08-557-146-2

Query Match 100.0%; Score 47; DB 2; Length 253;  
Best Local Similarity 100.0%; Pred. No. 0.49;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLVNERWVL 9  
|||||||  
DB 58 VLVNERWVL 66

## RESULT 11

US-08-824-874-3

Sequence 3, Application US/08824874  
Patent No. 5962300  
GENERAL INFORMATION:  
APPLICANT: Hillman, Jennifer L.  
TITLE OF INVENTION: NOVEL KALLIKREIN  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/824,874  
FILING DATE: Filed Herewith  
CLASSIFICATION: 514

PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:

ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.

REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0252 US

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166

INFORMATION FOR SEQ ID NO. 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 253 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear

IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 532504

US-08-824-874-3

Query Match 100.0%; Score 47; DB 2; Length 253;  
Best Local Similarity 100.0%; Pred. No. 0.49;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLVNERWVL 9  
|||||||  
DB 58 VLVNERWVL 66

## RESULT 12

US-09-154-344-2

Sequence 2, Application US/09154344  
Patent No. 5981256

GENERAL INFORMATION:  
APPLICANT: Egelrud, Torbjorn

ADDRESSEE: Hanson, Lennart  
TITLE OF INVENTION: Recombinant Stratum Corneum Chymotryptic

NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: White & Case, Patent Department  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2787

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/154,344  
FILING DATE: 16-SEP-1998  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/557,146  
FILING DATE: 14-DEC-1995  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Sterner, Richard J.  
REGISTRATION NUMBER: 35,372  
REFERENCE/DOCKET NUMBER: 1103326-181  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 819-8783  
TELEFAX: (212) 354-8113  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 253 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-154-344-2

Query Match 100.0%; Score 47; DB 2; Length 253;  
Best Local Similarity 100.0%; Pred. No. 0.49;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLVNERWVL 9  
DB 58 VLVNERWVL 66

RESULT 13  
US-09-188-2  
Sequence 2, Application US/08930188  
Patent No. 6093397  
GENERAL INFORMATION:  
APPLICANT: Dixon, Eric P.  
APPLICANT: Johnstone, Edward M.  
APPLICANT: Little, Sheila P.  
TITLE OF INVENTION: AMYLOID PRECURSOR PROTEIN PROTEASE AND  
RELATED NUCLEIC ACIDS  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Eli Lilly and Company  
STREET: Lilly Corporate Center  
CITY: Indianapolis  
STATE: Indiana  
COUNTRY: United States of America  
ZIP: 46285  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/930,188  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/416,257  
FILING DATE: 04-APR-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Blalock, Donna K.  
REGISTRATION NUMBER: 38,082  
REFERENCE/DOCKET NUMBER: X9239  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 317-277-1090  
TELEFAX: 317-276-3861  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:

LENGTH: 253 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-930-188-2

Query Match 100.0%; Score 47; DB 3; Length 253;  
Best Local Similarity 100.0%; Pred. No. 0.49;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLVNERWVL 9  
DB 58 VLVNERWVL 66

RESULT 14  
US-09-210-084-3  
Sequence 3, Application US/09210084  
Patent No. 6197511  
GENERAL INFORMATION:  
APPLICANT: Hillman, Jennifer L.  
APPLICANT: Lal, Preeti  
TITLE OF INVENTION: NOVEL KALLIKREIN  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/210,084  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/824,874  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0252 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 253 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 532504  
US-09-210-084-3

Query Match 100.0%; Score 47; DB 3; Length 253;  
Best Local Similarity 100.0%; Pred. No. 0.49;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLVNERWVL 9  
DB 58 VLVNERWVL 66

RESULT 15  
US-09-764-762-3



Sequence 3, Application US/09764762  
Patent No. 6472195  
GENERAL INFORMATION:  
APPLICANT: Hillman, Jennifer L.  
Lal, Preeti  
TITLE OF INVENTION: NOVEL KALLIKREIN  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09764,762  
FILING DATE: 16-Jan-2001  
CLASSIFICATION: <unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/210,084  
FILING DATE: <unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0252 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 253 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 532504  
SEQUENCE DESCRIPTION: SEQ ID NO: 3:  
US-09-764-762-3

Query Match 100.0%; Score 47; DB 4; Length 253;  
Best Local Similarity 100.0%; Pred. No. 0.49;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLVNERWVL 9  
Db 58 VLVNERWVL 66

Search completed: March 1, 2004, 17:38:24  
JOB time : 12.8869 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

protein - protein search, using sw model

on: March 1, 2004, 17:16:55 ; Search time 45.5556 Seconds  
(without alignments)  
55.820 Million cell updates/sec

File: US-09-905-083-34

Perfect score: 47

Sequence: 1 VLVNERWVL 9

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: A\_Geneseq\_29Jan04.\*

1: Geneseq1980s.\*  
2: Geneseq1990s.\*  
3: Geneseq2000s.\*  
4: Geneseq2001s.\*  
5: Geneseq2002s.\*  
6: Geneseq2003as.\*  
7: Geneseq2003bs.\*  
8: Geneseq2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	47	100.0	9	4	Aae08239 Human str
2	47	100.0	97	6	Ada05740 Human NOV
3	47	100.0	136	4	Adg23378 Novel hum
4	47	100.0	181	6	Ada05738 Human NOV
5	47	100.0	198	6	Ada05736 Human NOV
6	47	100.0	224	6	Ada05744 Human NOV
7	47	100.0	225	4	Aab98502 Human Str
8	47	100.0	247	6	Ada05742 Human NOV
9	47	100.0	250	6	Ada05732 Human NOV
10	47	100.0	252	6	Ada05734 Human NOV
11	47	100.0	253	2	Aar67888 Human str
12	47	100.0	253	2	Aaw05383 Human amy
13	47	100.0	253	5	Abb84421 Human SCC
14	47	100.0	253	5	Abb84406 Human SCC
15	47	100.0	253	5	Aau07440 Amino aci
16	47	100.0	253	6	Abu07471 Protein d
17	47	100.0	253	6	Abu07471 Protein d
18	47	100.0	253	6	Abf58471 Human str
19	47	100.0	253	7	Adh80484 Ovarian c
20	47	100.0	257	3	Aab21326 Human HSC
21	43	91.5	9	4	Aae08245 Human str
22	43	91.5	9	4	Aae08255 Human str
23	40	85.1	243	5	Abb84419 Bovine SC
24	40	85.1	249	5	Abb84420 Porcine S
25	39	83.0	212	2	Aay28590 Human Fac

26 39 83.0 228 7 AAE39994 Human adi  
27 39 83.0 250 2 AAR05421 Human adi  
28 39 83.0 253 7 AAE39992 Human adi  
29 38 80.9 64 2 AAW01196 Serine pr  
30 38 80.9 64 4 AAB50593 Flea seri  
31 38 80.9 81 6 ABG75788 Serine pr  
32 38 80.9 223 2 AAW01165  
33 38 80.9 224 4 AAB50561 Flea seri  
34 38 80.9 222 2 AAW64261 Kallikrei  
35 38 80.9 260 2 AAW10694 Human rec  
36 38 80.9 260 2 AAW12393 Mouse neu  
37 38 80.9 260 5 ABB57219 Mouse isc  
38 37 78.7 233 2 AAW46773 Amino aci  
39 37 78.7 233 4 AAG79000 Mamushi f  
40 37 78.7 237 2 AAW41955  
41 37 78.7 243 4 ABB60343 Drosophil  
42 37 78.7 247 6 ADA50480 Human pro  
43 37 78.7 258 2 AAW41954 Flea seri  
44 37 78.7 258 2 AAW41953 Flea seri  
45 37 78.7 271 7 ADB97555 Human MTS

## ALIGNMENTS

RESULT 1  
AAE08239  
ID AAE08239 standard; peptide; 9 AA.  
AC AAE08239;  
XX  
DT 01-NOV-2001 (first entry)  
XX  
DE Human stratum corneum chymotrypsin enzyme peptide #3 (residues 58-66).  
XX  
KW Stratum corneum chymotrypsin enzyme; SCCE; cytostatic; vaccine; tumour;  
KW cancer; ovarian; breast; lung; colon; prostate; carcinoma; human;  
KW antisense therapy; malignant hyperplasia.  
XX  
OS Homo sapiens.  
XX  
PN WO200159158-A1.  
XX  
PD 16-AUG-2001.  
XX  
PF 07-FEB-2001; 2001WO-US003977.  
XX  
PR 11-FEB-2000; 2000US-00502600.  
XX  
PA (UYAR-) UNIV ARKANSAS.  
XX  
PI O'brien TJ;  
XX  
DR WPI; 2001-514676/56.  
XX  
PT Diagnosing cancer comprises detecting stratum corneum chymotrypsin enzyme.  
XX  
PS Claim 25; Page 103; 127pp; English.  
XX  
CC The invention relates to diagnosing cancer especially ovarian cancer, by screening for stratum corneum chymotrypsin enzyme (SCCE). Proteases are considered to be an integral part of tumour growth and metastasis, and therefore, markers indicative of their presence or absence are useful for the diagnosis of cancer. The method is useful for diagnosing cancer, CC carcinoma and malignant hyperplasia. The SCCE oligonucleotide may be used to treat a cancer selected from ovarian, breast, lung, colon, prostate and other cancers in which SCCE is overexpressed. The present sequence is human SCCE peptide  
XX  
SQ Sequence 9 AA;  
Query Match 100.0%; Score 47; DB 4; Length 9;

Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

/ 1 VLVNERWVL 9  
| | | | |  
b 1 VLVNERWVL 9

RESULT 2  
D ADR05740 standard; protein; 97 AA.  
X C ADA05740;  
X X 06-NOV-2003 (first entry)  
X X Human NOV18e protein SEQ ID NO:100.  
X X human; NOVX; antidiabetic; anorectic; antibacterial; virucide;  
X X immunomodulator; cytostatic; neurotropic; neuroprotective;  
X X antiparkinsonian; antilipemic; gene therapy; human disease;  
X X metabolic disorder; diabetes; obesity; infection; cachexia; cancer;  
X X neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;  
X X immune disorder; haematopoietic disorder; dyslipidaemia.  
X X Homo sapiens.  
X X WO2003029424-A2.  
X X  
X X 10-APR-2003.  
X X  
X X 02-OCT-2002; 2002WO-US031373.  
X X  
X X 02-OCT-2001; 2001US-0326483P.  
X X 05-OCT-2001; 2001US-0327435P.  
X X 05-OCT-2001; 2001US-0327449P.  
X X 09-OCT-2001; 2001US-0327917P.  
X X 09-OCT-2001; 2001US-0328029P.  
X X 09-OCT-2001; 2001US-0328044P.  
X X 09-OCT-2001; 2001US-0328056P.  
X X 12-OCT-2001; 2001US-0328849P.  
X X 15-OCT-2001; 2001US-0329414P.  
X X 17-OCT-2001; 2001US-0330142P.  
X X 18-OCT-2001; 2001US-0330309P.  
X X 22-OCT-2001; 2001US-0341058P.  
X X 24-OCT-2001; 2001US-0339266P.  
X X 24-OCT-2001; 2001US-0343629P.  
X X 29-OCT-2001; 2001US-0349575P.  
X X 01-NOV-2001; 2001US-0346357P.  
X X 17-APR-2002; 2002US-0373260P.  
X X 19-APR-2002; 2002US-0373815P.  
X X 19-APR-2002; 2002US-0373817P.  
X X 19-APR-2002; 2002US-0373826P.  
X X 19-APR-2002; 2002US-0373884P.  
X X 22-APR-2002; 2002US-0374977P.  
X X 16-MAY-2002; 2002US-0381037P.  
X X 16-MAY-2002; 2002US-0381038P.  
X X 16-MAY-2002; 2002US-0381042P.  
X X 17-MAY-2002; 2002US-0381642P.  
X X 28-MAY-2002; 2002US-0383656P.  
X X 29-MAY-2002; 2002US-0383831P.  
X X 25-JUN-2002; 2002US-0391335P.  
X X 01-OCT-2002; 2002US-00262511.  
X X  
X X (CURA-) CURAGEN CORP.  
X X  
X X Smithson G, Millet I, Feyman JA, Kekuda R, Ju J, Li L, Guo X;  
X X Patturajan M, Spytek KA, Edinger SR, Ellerman K, Malyankar UM;  
X X Ort T, Gorman L, Zerkhusen BD, Anderson DW, Zhong M, Catterton E;  
X X Ji W, Miller CB, Rastelli L, Stone DJ, Pena CE, Shenoy SG;  
X X Shinkets RA, Rothenberg ME, Leach MD, Agee ML, Berghs C, Dipippo VA;  
X X Eisen AJ, Gangolli EA, Rieger DK, Spaderna SK;

DR WPI; 2003-381626/36.  
DR N-PSDB; ADA05739.  
XX  
XX New NOVX polypeptides and nucleic acids, useful for diagnosing,  
PT preventing or treating NOVX-associated disorders, e.g. diabetes, obesity,  
PT cancer or dyslipidemia, and in chromosome mapping, tissue typing or  
PT pharmacogenomics.  
XX  
XX Claim 1; Page 171; 586pp; English.  
PS  
XX The present invention describes NOVX proteins, where X can be 1 to 55  
CC (e.g. NOV1). Also described: (1) a composition comprising a polypeptide  
CC described above and a carrier; (2) a kit comprising, in one or more  
CC containers, the composition described above; (3) an isolated nucleic acid  
CC molecule which encodes a NOVX protein of the invention; (4) a vector  
CC comprising the nucleic acid molecule described above; (5) a cell  
CC comprising the above vector; (6) an antibody that immunospecifically  
CC binds to the polypeptide described above; (7) methods for determining the  
CC presence or amount of the above polypeptide or nucleic acid molecule in a  
CC sample; (8) methods for determining the presence of or predisposition to  
CC a disease associated with altered levels of expression of the above  
CC polypeptide or nucleic acid molecule in a first mammalian subject; (9) a  
CC method of identifying an agent that binds to the polypeptide described  
CC above; (10) a method for identifying a potential therapeutic agent for  
CC use in treating a pathology that is related to an aberrant expression or  
CC aberrant physiological interactions of the polypeptide; (11) a method of  
CC screening for a modulator of activity or of latency or predisposition to  
CC a pathology associated with the polypeptide; (12) a method for modulating  
CC the activity of the polypeptide associated with the above polypeptide in a  
CC mammal; and (14) a method for producing the above polypeptide. NOVX  
CC sequences have antidiabetic, anorectic, antibacterial, virucide  
CC immunomodulator, cytostatic, neurotropic, neuroprotective, antiparkinsonian  
CC and antilipemic activities, and can be used in gene therapy. The  
CC polypeptide is useful in manufacturing a medicament for treating a  
CC syndrome associated with a human disease. The polypeptide or the nucleic  
CC acid molecule may be used to diagnose, treat or prevent metabolic  
CC disorders such as diabetes or obesity, infections, cachexia, cancer,  
CC neurodegenerative disorders such as Alzheimer's disease or Parkinson's  
CC disease, immune disorders, haematopoietic disorders and various  
CC dyslipidaemias. The nucleic acids can also be used as hybridisation  
CC probes, in chromosome mapping, tissue typing, preventive medicine and  
CC pharmacogenomics. The present sequence represents a human NOVX from the  
CC present invention.  
XX

XX Sequence 97 AA;

Query Match 100.0%; Score 47; DB 6; Length 97;

Best Local Similarity 100.0%; Pred. No. 0.87;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VLVNERWVL 9

Db 39 VLVNERWVL 47

RESULT 3

ABG23378

ID ABG23378 standard; protein; 136 AA.

XX

AC ABG23378;

XX

DT 18-FEB-2002 (first entry)

XX

DE Novel human diagnostic protein #23369.

XX

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder.

XX

OS Homo sapiens.

XX

PN WO200175067-A2.

XX

) 11-OCT-2001.  
 X 30-MAR-2001; 2001WO-US008631.  
 X 31-MAR-2000; 2000US-00540217.  
 X 23-AUG-2000; 2000US-00649167.  
 X (HYSE-) HYSEQ INC.  
 X Dmanac RT, Liu C, Tang YT;  
 X WPI; 2001-639362/73.  
 X N-PSDB; AAS87565.  
 X New isolated polynucleotide and encoded polypeptides, useful in  
 X diagnostics, forensics, gene mapping, identification of mutations  
 X responsible for genetic disorders or other traits and to assess  
 X biodiversity.  
 X Claim 20; SEQ ID NO 5377; 103pp; English.  
 X The invention relates to isolated polynucleotide (I) and polypeptide (II)  
 X sequences. (I) is useful as hybridisation probes, polymerase chain  
 X reaction (PCR) primers, oligomers, and for chromosome and gene mapping.  
 X and in recombinant production of (II). The polynucleotides are also used  
 X in diagnostics as expressed sequence tags for identifying expressed  
 X genes. (I) is useful in gene therapy techniques to restore normal  
 X activity of (II) or to treat disease states involving (II). (II) is  
 X useful for generating antibodies against it, detecting or quantitating a  
 X polypeptide in tissue, as molecular weight markers and as a food  
 X supplement. (II) and its binding partners are useful in medical imaging  
 X of sites expressing (II). (I) and (II) are useful for treating disorders  
 X involving aberrant protein expression or biological activity. The  
 X polypeptide and polynucleotide sequences have applications in  
 X diagnostics, forensics, gene mapping, identification of mutations  
 X responsible for genetic disorders or other traits to assess biodiversity  
 X and to produce other types of data and products dependent on DNA and  
 X amino acid sequences. ASG00010-ASG30377 represent novel human diagnostic  
 X amino acid sequences of the invention. Note: The sequence data for this  
 X patent did not appear in the printed specification, but was obtained in  
 X electronic format directly from WIPO at  
 X ftp.wipo.int/pub/published\_pct\_sequences  
 X Sequence 136 AA;  
 X  
 X Query Match 100.0%; Score 47; DB 4; Length 136;  
 X Best Local Similarity 100.0%; Pred. No. 1.3;  
 X Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 X  
 X QY 1 VLVNERWVL 9  
 X 64 VLVNERWVL 72  
 X  
 X RESULT 4  
 X ADA05738  
 X ID ADA05738 standard; protein; 181 AA.  
 X AC ADA05738;  
 X DT 06-NOV-2003 (first entry)  
 X DE Human NOV18d protein SEQ ID NO:98.  
 X KW human; NOX; antidiabetic; anorectic; antibacterial; virucide;  
 X KW immunomodulator; cytostatic; nootropic; neuroprotective;  
 X KW antiparkinsonian; antilipemic; gene therapy; human disease;  
 X KW metabolic disorder; diabetes; obesity; infection; cachexia; cancer;  
 X KW neurodegenerative disorder; Alzheimer's disease; parkinson's disease;  
 X KW immune disorder; haematopoietic disorder; dyslipidaemia.  
 X OS Homo sapiens.  
 X

PN WO2003029424-A2.  
 XX 10-APR-2003.  
 PD 02-OCT-2002; 2002WO-US031373.  
 XX 02-OCT-2001; 2001US-0326483P.  
 PR 03-OCT-2001; 2001US-0327435P.  
 PR 05-OCT-2001; 2001US-0327449P.  
 PR 09-OCT-2001; 2001US-0327917P.  
 PR 09-OCT-2001; 2001US-0328029P.  
 PR 09-OCT-2001; 2001US-0328044P.  
 PR 09-OCT-2001; 2001US-0328056P.  
 PR 12-OCT-2001; 2001US-0328849P.  
 PR 15-OCT-2001; 2001US-0329414P.  
 PR 17-OCT-2001; 2001US-0330142P.  
 PR 18-OCT-2001; 2001US-0330309P.  
 PR 22-OCT-2001; 2001US-0340588P.  
 PR 24-OCT-2001; 2001US-0339266P.  
 PR 24-OCT-2001; 2001US-0343629P.  
 PR 29-OCT-2001; 2001US-0349575P.  
 PR 01-NOV-2001; 2001US-0346357P.  
 PR 17-APR-2002; 2002US-0373260P.  
 PR 19-APR-2002; 2002US-0373815P.  
 PR 19-APR-2002; 2002US-0373817P.  
 PR 19-APR-2002; 2002US-0373826P.  
 PR 19-APR-2002; 2002US-0373884P.  
 PR 22-APR-2002; 2002US-0374977P.  
 PR 16-MAY-2002; 2002US-0381037P.  
 PR 16-MAY-2002; 2002US-0381038P.  
 PR 16-MAY-2002; 2002US-0381042P.  
 PR 17-MAY-2002; 2002US-0381642P.  
 PR 28-MAY-2002; 2002US-0383656P.  
 PR 29-MAY-2002; 2002US-0383831P.  
 PR 25-JUN-2002; 2002US-0391335P.  
 PR 01-OCT-2002; 2002US-00362511.  
 XX (CURA-) CURAGEN CORP.  
 XX Smithson G, Millet I, Peyman JA, Kekuda R, Ju J, Li L, Guo X;  
 PI Patturajan M, Spytek KA, Edinger SR, Ellerman K, Malyankar DM;  
 PI Ort T, Gorman L, Zerhusen BD, Anderson DW, Zhong M, Catterton E;  
 PI Ji W, Miller CE, Rastelli L, Stone DJ, Pena CE, Shenoy SG;  
 PI Shimkets RA, Rothenberg ME, Leach MD, Agee ML, Berghs C, Dipippo VA;  
 PI Eisen AJ, Gangolli EA, Rieger DK, Spaderna SK;  
 XX WPI; 2003-381626/36.  
 DR N-PSDB; ADA05737.  
 XX  
 XX New NOX polypeptides and nucleic acids, useful for diagnosing,  
 PT preventing or treating NOX-associated disorders, e.g. diabetes, obesity,  
 PT cancer or dyslipidemia, and in chromosome mapping, tissue typing or  
 PT pharmacogenomics.  
 XX  
 XX Claim 1; Page 171; 586pp; English.  
 XX The present invention describes NOX proteins, where X can be 1 to 55  
 CC (e.g. NOX1). Also described: (1) a composition comprising a polypeptide  
 CC described above and a carrier; (2) a kit comprising, in one or more  
 CC containers, the composition described above; (3) an isolated nucleic acid  
 CC molecule which encodes a NOX protein of the invention; (4) a vector  
 CC comprising the nucleic acid molecule described above; (5) a cell  
 CC comprising the above vector; (6) an antibody that immunospecifically  
 CC binds to the polypeptide described above; (7) methods for determining the  
 CC presence or amount of the above polypeptide or nucleic acid molecule in a  
 CC sample; (8) methods for determining the presence of or predisposition to  
 CC a disease associated with altered levels of expression of the above  
 CC polypeptide or nucleic acid molecule in a first mammalian subject; (9) a  
 CC method of identifying an agent that binds to the polypeptide described  
 CC above; (10) a method for identifying a potential therapeutic agent for  
 CC use in treating a pathology that is related to an aberrant expression or  
 CC aberrant physiological interactions of the polypeptide; (11) a method of  
 CC screening for a modulator of activity or of latency or predisposition to

C a pathology associated with the polypeptide; (12) a method for modulating the activity of the polypeptide described above; (13) methods of treating or preventing a pathology associated with the above polypeptide in a mammal; and (14) a method for producing the above polypeptide. NOVX sequences have antidiabetic, anorectic, antibacterial, virucide, immunomodulator, cytostatic, neurotropic, neuroprotective, antiparkinsonian and antilipaeamic activities, and can be used in gene therapy. The polypeptide is useful in manufacturing a medicament for treating a syndrome associated with a human disease. The polypeptide or the nucleic acid molecule may be used to diagnose, treat or prevent metabolic disorders such as diabetes or obesity, infections, cachexia, cancer, neurodegenerative disorders such as Alzheimer's disease or Parkinson's disease, immune disorders, haematopoietic disorders and various dyslipidaemias. The nucleic acids can also be used as hybridisation probes, in chromosome mapping, tissue typing, preventive medicine and pharmacogenomics. The present sequence represents a human NOVX from the present invention.

XX Sequence 181 AA;

Query Match 100.0%; Score 47; DB 6; Length 181;  
Best Local Similarity 100.0%; Pred. No. 1.7;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLVNRRWL 9  
|||||  
Db 39 VLVNRRWL 47

RESULT 5  
ADA05736  
ID ADA05736 standard; protein; 198 AA.

XX ADA05736;

DT 06-NOV-2003 (first entry)

XX Human NOV18c protein SEQ ID NO:96.

XX human; NOVX; antidiabetic; anorectic; antibacterial; virucide;  
KW immunomodulator; cytostatic; neurotropic; neuroprotective;  
KW antiparkinsonian; antilipaeamic; gene therapy; human disease;  
KW metabolic disorder; diabetes; obesity; infection; cachexia; cancer;  
KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;  
KW immune disorder; haematopoietic disorder; dyslipidaemia.

XX Homo sapiens.

XX WO2003029424-A2.

XX 10-APR-2003.

XX 02-OCT-2002; 2002WO-US031373.

XX 02-OCT-2001; 2001US-0326483P.

XX 05-OCT-2001; 2001US-0327435P.

XX 09-OCT-2001; 2001US-0327449P.

XX 09-OCT-2001; 2001US-0327917P.

XX 09-OCT-2001; 2001US-0328029P.

XX 09-OCT-2001; 2001US-0328044P.

XX 09-OCT-2001; 2001US-0328056P.

XX 12-OCT-2001; 2001US-0328849P.

XX 15-OCT-2001; 2001US-0329414P.

XX 17-OCT-2001; 2001US-0330142P.

XX 18-OCT-2001; 2001US-0330309P.

XX 22-OCT-2001; 2001US-0341058P.

XX 24-OCT-2001; 2001US-0339266P.

XX 24-OCT-2001; 2001US-0343629P.

XX 29-OCT-2001; 2001US-0349575P.

XX 01-NOV-2001; 2001US-0346357P.

XX 17-APR-2002; 2002US-0373260P.

XX 19-APR-2002; 2002US-0373815P.

XX 19-APR-2002; 2002US-0373817P.

19-APR-2002; 2002US-0373826P.  
PR 19-APR-2002; 2002US-0373884P.  
PR 22-APR-2002; 2002US-0374977P.  
PR 16-MAY-2002; 2002US-0381037P.  
PR 16-MAY-2002; 2002US-0381038P.  
PR 16-MAY-2002; 2002US-0381042P.  
PR 17-MAY-2002; 2002US-0381642P.  
PR 28-MAY-2002; 2002US-0383656P.  
PR 29-MAY-2002; 2002US-0383631P.  
PR 25-JUN-2002; 2002US-0391333P.  
PR 01-OCT-2002; 2002US-00262511.  
XX (CURA-) CURAGEN CORP.  
PI Smithe G, Millet I, Peyman JA, Kekuda R, Ju J, Li L, Guo X;  
PI Patturajan M, Spytek KA, Edinger SR, Ellerman K, Malyankar UM;  
PI Ott T, Gorman L, Zerhusen BD, Anderson DW, Zhong M, Cattarton E;  
PI Ji W, Miller CE, Rastelli L, Stone DJ, Pena CE, Shenoy SG;  
PI Shimkets RA, Rothenberg MS, Leach MD, Adee ML, Berghs C, Dippio VA;  
PI Eisen AJ, Gangolli EA, Rieger DK, Spaderna SK;  
XX WPI; 2003-381626/36.  
DR N-PSDB; ADA05735.

New NOVX polypeptides and nucleic acids, useful for diagnosing, preventing or treating NOVX-associated disorders, e.g. diabetes, obesity, cancer or dyslipidemia, and in chromosome mapping, tissue typing or pharmacogenomics.

Claim 1; Page 170; 586pp; English.

The present invention describes NOVX proteins, where X can be 1 to 55 (e.g. NOV1). Also described: (1) a composition comprising a polypeptide described above and a carrier; (2) a kit comprising, in one or more containers, the composition described above; (3) an isolated nucleic acid molecule which encodes a NOVX protein of the invention; (4) a vector comprising the nucleic acid molecule described above; (5) a cell comprising the above vector; (6) an antibody that immunospecifically binds to the polypeptide described above; (7) methods for determining the presence or amount of the above polypeptide or nucleic acid molecule in a sample; (8) methods for determining the presence of or predisposition to a disease associated with altered levels of expression of the above polypeptide or nucleic acid molecule in a first mammalian subject; (9) a method of identifying an agent that binds to the polypeptide described above; (10) a method for identifying a potential therapeutic agent for use in treating a pathology that is related to an aberrant expression or aberrant physiological interactions of the polypeptide; (11) a method of screening for a modulator of activity of the polypeptide; (12) a method of screening for a pathology associated with the polypeptide; (13) methods of treating or preventing a pathology associated with the above polypeptide in a mammal; and (14) a method for producing the above polypeptide. NOVX sequences have antidiabetic, anorectic, antibacterial, virucide, immunomodulator, cytostatic, neurotropic, neuroprotective, antiparkinsonian and antilipaeamic activities, and can be used in gene therapy. The polypeptide is useful in manufacturing a medicament for treating a syndrome associated with a human disease. The polypeptide or the nucleic acid molecule may be used to diagnose, treat or prevent metabolic disorders such as diabetes or obesity, infections, cachexia, cancer, neurodegenerative disorders such as Alzheimer's disease or Parkinson's disease, immune disorders, haematopoietic disorders and various dyslipidaemias. The nucleic acids can also be used as hybridisation probes, in chromosome mapping, tissue typing, preventive medicine and pharmacogenomics. The present sequence represents a human NOVX from the present invention.

XX Sequence 198 AA;

Query Match 100.0%; Score 47; DB 6; Length 198;  
Best Local Similarity 100.0%; Pred. No. 1.9;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLVNRRWL 9

|||||  
58 VLVNERWVL 66

RESULT 6  
JA05744  
J ADA05744 standard; protein; 224 AA.

K ADA05744;

L 06-NOV-2003 (first entry)

M Human NOV18g protein SEQ ID NO:104.

N human; NOVX; antidiabetic; anorectic; antibacterial; virucide;  
O immunomodulator; cytosolic; nontropic; neuroprotective;  
P antiparkinsonian; antipalemic; gene therapy; human disease;  
Q metabolic disorder; diabetes; obesity; infection; cachexia; cancer;  
R neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;  
S immune disorder; haematopoietic disorder; dyslipidaemia.

T Homo sapiens.

U WO2003029424-A2.

V 10-APR-2003.

W 02-OCT-2002; 2002WO-US031373.

X 02-OCT-2001; 2001US-0326483P.

Y 05-OCT-2001; 2001US-0327435P.

Z 05-OCT-2001; 2001US-0327449P.

[ 09-OCT-2001; 2001US-0327917P.

^ 09-OCT-2001; 2001US-0328029P.

\_ 09-OCT-2001; 2001US-0328044P.

` 09-OCT-2001; 2001US-0328056P.

~ 12-OCT-2001; 2001US-0328849P.

1 15-OCT-2001; 2001US-0329414P.

2 17-OCT-2001; 2001US-0330142P.

3 18-OCT-2001; 2001US-0330309P.

4 22-OCT-2001; 2001US-0341058P.

5 24-OCT-2001; 2001US-0339266P.

6 24-OCT-2001; 2001US-0343629P.

7 29-OCT-2001; 2001US-0349575P.

8 01-NOV-2001; 2001US-0346337P.

9 17-APR-2002; 2002US-0373260P.

0 19-APR-2002; 2002US-0373815P.

1 19-APR-2002; 2002US-0373817P.

2 19-APR-2002; 2002US-0373856P.

3 19-APR-2002; 2002US-0373884P.

4 22-APR-2002; 2002US-0374977P.

5 16-MAY-2002; 2002US-0381037P.

6 16-MAY-2002; 2002US-0381038P.

7 17-MAY-2002; 2002US-0381042P.

8 28-MAY-2002; 2002US-0381642P.

9 29-MAY-2002; 2002US-0383656P.

0 29-MAY-2002; 2002US-0383831P.

1 25-JUN-2002; 2002US-0391335P.

2 01-OCT-2002; 2002US-00262511.

(CURA-) CURAGEN CORP.

Smithson G, Millet I, Peyman JA, Kekuda R, Ju J, Li L, Guo X;  
Patturajan M, Sytek KA, Edinger SR, Ellerman K, Malyankar UM;  
Ort T, Gorman L, Zerhusen BD, Anderson DW, Zhong M, Catterton E;  
Ji W, Miller CE, Rastelli L, Stone DJ, Pena CE, Shenoy SG;  
Shinkets RA, Rothenberg WB, Leach MD, Agee ML, Berghs C, Dipippo VA;  
Eisen AJ, Gangolli EA, Rieger DK, Spaderna SK;  
WPI; 2003-381626/36.  
N-PSDB; ADA05743.

New NOVX polypeptides and nucleic acids, useful for diagnosing,

PT preventing or treating NOVX-associated disorders, e.g. diabetes, obesity,  
PT cancer or dyslipidemia, and in chromosome mapping, tissue typing or  
PT pharmacogenomics.

PS Claim 1; Page 172; 586pp; English.

XX The present invention describes NOVX proteins, where X can be 1 to 55  
CC (e.g. NOV1). Also described: (1) a composition comprising a polypeptide  
CC described above and a carrier; (2) a kit comprising, in one or more  
CC containers, the composition described above; (3) an isolated nucleic acid  
CC molecule which encodes a NOVX protein of the invention; (4) a vector  
CC comprising the nucleic acid molecule described above; (5) a cell  
CC comprising the above vector; (6) an antibody that immunospecifically  
CC binds to the polypeptide described above; (7) methods for determining the  
CC presence or amount of the above polypeptide or nucleic acid molecule in a  
CC sample; (8) methods for determining the presence of or predisposition to  
CC a disease associated with altered levels of expression of the above  
CC polypeptide or nucleic acid molecule in a first mammalian subject; (9) a  
CC method of identifying an agent that binds to the polypeptide described  
CC above; (10) a method for identifying a potential therapeutic agent for  
CC use in treating a pathology that is related to an aberrant expression or  
CC aberrant physiological interactions of the polypeptide; (11) a method of  
CC screening for a modulator of activity or of latency or predisposition to  
CC a pathology associated with the polypeptide; (12) a method for modulating  
CC the activity of the polypeptide described above; (13) methods of treating  
CC or preventing a pathology associated with the above polypeptide in a  
CC mammal; and (14) a method for producing the above polypeptide. NOVX  
CC sequences have antidiabetic, anorectic, antibacterial, virucide,  
CC immunomodulator, cytosolic, nontropic, neuroprotective, antiparkinsonian  
CC and antipalemic activities, and can be used in gene therapy. The  
CC polypeptide is useful in manufacturing a medicament for treating a  
CC syndrome associated with a human disease. The polypeptide or the nucleic  
CC acid molecule may be used to diagnose, treat or prevent metabolic  
CC disorders such as diabetes or obesity, infections, cachexia, cancer,  
CC neurodegenerative disorders such as Alzheimer's disease or Parkinson's  
CC disease, immune disorders, haematopoietic disorders and various  
CC dyslipidaemias. The nucleic acids can also be used as hybridisation  
CC probes, in chromosome mapping, tissue typing, preventive medicine and  
CC pharmacogenomics. The present sequence represents a human NOVX from the  
CC present invention.

XX Sequence 224 AA;

Query Match 100.0%; Score 47; DB 6; Length 224;

Best Local Similarity 100.0%; Pred. NO. 2.2;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLVNERWVL 9

Db 44 VLVNERWVL 52

RESULT 7

AAB98502

ID AAB98502 standard; protein; 225 AA.

XX AAB98502;

XX 03-AUG-2001 (first entry)

XX Human Stratum Corneum Chymotryptic Enzyme, SCCE, catalytic domain.

XX Human; TAG-15; cytostatic; vaccine; ovarian tumour; cancer;

XX tumour antigen-derived gene 15; serine protease;

XX Stratum Corneum Chymotryptic Enzyme; SCCE.

OS Homo sapiens.

XX WC200129056-A1.

XX 26-APR-2001.

XX 20-OCT-2000; 2000WO-US029095.

20-OCT-1999; 99US-00421213.  
 (UYAR-) UNIV ARKANSAS.  
 O'Brien TJ, Tanimoto H;  
 WPI; 2001-381031/40.  
 Novel extracellular serine protease, termed tumor antigen-derived gene 15  
 protein overexpressed in carcinomas and DNA encoding it, for diagnosis,  
 treatment, prevention of cancer, particularly breast, ovarian cancer.  
 Example 10; Fig 1; 130pp; English.  
 The present invention relates to human tumor antigen-derived gene 15  
 (TAGD-15) protein and coding sequence (see AAH23601 and AB985001). TAGD-  
 15 is an extracellular serine protease. It was found that TAGD-15 is over-  
 expressed in ovarian tumors. TAGD-15 protein or its fragments of 9-20  
 residues that lack TAGD-15 protease activity are useful for vaccinating  
 an individual against TAGD-15, having, suspected of having or at risk of  
 getting cancer. Furthermore, the TAGD-15 gene can be used as a diagnostic  
 or therapeutic target in cancer. The present sequence was used in a  
 sequence homology alignment with the catalytic domain of TAGD-15  
 Sequence 225 AA;  
 Query Match 100.0%; Score 47; DB 4; Length 225;  
 Best Local Similarity 100.0%; Pred. No. 2.2; Indels 0; Gaps 0;  
 Matches 9; Conservative 0; Mismatches 0;  
 QY 1 VLVNERWVL 9  
 Db 30 VLVNERWVL 38  
 RESULT 8  
 ADA05742  
 ID ADA05742 standard; protein; 247 AA.  
 AC ADA05742;  
 DT 06-NOV-2003 (first entry)  
 DE Human NOV18f protein SEQ ID NO:102.  
 KW human; NOVX; antidiabetic; anorectic; antibacterial; virucide;  
 KW immunomodulator; cytostatic; nootropic; neuroprotective;  
 KW antiparkinsonian; antilipemic; gene therapy; human disease;  
 KW metabolic disorder; diabetes; obesity; infection; cachexia; cancer;  
 KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;  
 KW immune disorder; haematopoietic disorder; dyslipidaemia.  
 OS Homo sapiens.  
 PN WO2003029424-A2.  
 PD 10-APR-2003.  
 XX 02-OCT-2002; 2002WO-US031373.  
 XX 02-OCT-2001; 2001US-0326483P.  
 PR 05-OCT-2001; 2001US-0327435P.  
 PR 05-OCT-2001; 2001US-0327449P.  
 PR 09-OCT-2001; 2001US-0327917P.  
 PR 09-OCT-2001; 2001US-0328029P.  
 PR 09-OCT-2001; 2001US-0328044P.  
 PR 09-OCT-2001; 2001US-0328056P.  
 PR 12-OCT-2001; 2001US-0328849P.  
 PR 15-OCT-2001; 2001US-0329414P.  
 PR 17-OCT-2001; 2001US-0330142P.  
 PR 18-OCT-2001; 2001US-0330309P.  
 PR 22-OCT-2001; 2001US-0341058P.

24-OCT-2001; 2001US-0339266P.  
 PR 24-OCT-2001; 2001US-0343629P.  
 PR 29-OCT-2001; 2001US-0349575P.  
 PR 01-NOV-2001; 2001US-0346357P.  
 PR 17-APR-2002; 2002US-0373260P.  
 PR 19-APR-2002; 2002US-0373815P.  
 PR 19-APR-2002; 2002US-0373815P.  
 PR 19-APR-2002; 2002US-0373815P.  
 PR 19-APR-2002; 2002US-0373826P.  
 PR 19-APR-2002; 2002US-0373884P.  
 PR 22-APR-2002; 2002US-0374977P.  
 PR 16-MAY-2002; 2002US-0381037P.  
 PR 16-MAY-2002; 2002US-0381038P.  
 PR 16-MAY-2002; 2002US-0381042P.  
 PR 17-MAY-2002; 2002US-0381642P.  
 PR 28-MAY-2002; 2002US-0383656P.  
 PR 29-MAY-2002; 2002US-0383831P.  
 PR 25-JUN-2002; 2002US-0391335P.  
 PR 01-OCT-2002; 2002US-00262511.  
 XX (CURA-) CURAGEN CORP.  
 PA Smithson G, Millet I, Peyman JA, Kekuda R, Ju J, Li L, Guo X;  
 PI Pattarajan M, Spytek KA, Edinger SR, Eilerman K, Malyanar UW;  
 PI Ott T, Gorman L, Zerhusen BD, Anderson DW, Zhong M, Catterton B;  
 PI Ji W, Miller CE, Rastelli L, Stone DJ, Pena CEA, Shenoy SG;  
 PI Shimkets RA, Rothenberg ME, Leach MD, Agee ML, Berghs C, Dipippo VA;  
 PI Eisen AJ, Gangolli EA, Rieger DK, Spaderna SK;  
 PI WPI; 2003-381626/36.  
 DR N-PSDB; ADA05741.  
 XX  
 PT New NOVX polypeptides and nucleic acids, useful for diagnosing,  
 PT preventing or treating NOVX-associated disorders, e.g. diabetes, obesity,  
 PT cancer or dyslipidemia, and in chromosome mapping, tissue typing or  
 PT pharmacogenomics.  
 XX  
 PS Claim 1; Page 172; 586pp; English.  
 XX  
 CC The present invention describes NOVX proteins, where X can be 1 to 55  
 CC (e.g. NOV1). Also described: (1) a composition comprising a polypeptide  
 CC described above and a carrier; (2) a kit comprising, in one or more  
 CC containers, the composition described above; (3) an isolated nucleic acid  
 CC molecule which encodes a NOVX protein of the invention; (4) a vector  
 CC comprising the nucleic acid molecule described above; (5) a cell  
 CC comprising the above vector; (6) an antibody that immunospecifically  
 CC binds to the polypeptide described above; (7) methods for determining the  
 CC presence or amount of the above polypeptide or nucleic acid molecule in a  
 CC sample; (8) methods for determining the presence of or predisposition to  
 CC a disease associated with altered levels of expression of the above  
 CC polypeptide or nucleic acid molecule in a first mammalian subject; (9) a  
 CC method of identifying an agent that binds to the polypeptide described  
 CC above; (10) a method for identifying a potential therapeutic agent for  
 CC use in treating a pathology that is related to an aberrant expression or  
 CC aberrant physiological interactions of the polypeptide; (11) a method of  
 CC screening for a modulator of activity or of latency or predisposition to  
 CC a pathology associated with the polypeptide; (12) a method for modulating  
 CC the activity of the polypeptide described above; (13) methods of treating  
 CC or preventing a pathology associated with the above polypeptide in a  
 CC mammal; and (14) a method for producing the above polypeptide. NOVX  
 CC sequences have antidiabetic, anorectic, antibacterial, virucide,  
 CC immunomodulator, cytostatic, nootropic, neuroprotective, antiparkinsonian  
 CC and antilipemic activities, and can be used in gene therapy. The  
 CC polypeptide is useful in manufacturing a medicament for treating a  
 CC syndrome associated with a human disease. The polypeptide or the nucleic  
 CC acid molecule may be used to diagnose, treat or prevent metabolic  
 CC disorders such as diabetes or obesity, infections, cachexia, cancer,  
 CC neurodegenerative disorders such as Alzheimer's disease or Parkinson's  
 CC disease, immune disorders, haematopoietic disorders and various  
 CC dyslipidaemias. The nucleic acids can also be used as hybridisation  
 CC probes, in chromosome mapping, tissue typing, preventive medicine and  
 CC pharmacogenomics. The present sequence represents a human NOVX from the  
 CC present invention.

2 Sequence 247 AA;  
 Query Match 100.0%; Score 47; DB 6; Length 247;  
 Best Local Similarity 100.0%; Pred. No. 2.5; Indels 0; Gaps 0;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 VLVNERWVL 9  
 b 44 VLVNERWVL 52

RESULT 9  
 ADA05732  
 ID ADA05732 standard; protein; 250 AA.  
 AC ADA05732;  
 DT 06-NOV-2003 (first entry)  
 DE Human NOV18a protein SEQ ID NO:92.  
 KW human; NOVX; antidiabetic; anorectic; antibacterial; virucide;  
 KW immunomodulator; cytostatic; nootropic; neuroprotective;  
 KW antiparkinsonian; antilipemic; gene therapy; human disease;  
 KW metabolic disorder; diabetes; obesity; infection; cachexia; cancer;  
 KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;  
 KW immune disorder; haematopoietic disorder; dyslipidaemia.  
 OS Homo sapiens.  
 XX  
 XX WO2003029424-A2.  
 XX  
 XX 10-APR-2003.  
 XX  
 XX 02-OCT-2002; 2002WO-US031373.  
 XX  
 PR 02-OCT-2001; 2001US-0326483P.  
 PR 03-OCT-2001; 2001US-0327435P.  
 PR 03-OCT-2001; 2001US-0327449P.  
 PR 09-OCT-2001; 2001US-0327917P.  
 PR 09-OCT-2001; 2001US-0328029P.  
 PR 09-OCT-2001; 2001US-0328044P.  
 PR 09-OCT-2001; 2001US-0328056P.  
 PR 12-OCT-2001; 2001US-0328849P.  
 PR 15-OCT-2001; 2001US-0329414P.  
 PR 17-OCT-2001; 2001US-0330142P.  
 PR 18-OCT-2001; 2001US-0330309P.  
 PR 22-OCT-2001; 2001US-0341058P.  
 PR 24-OCT-2001; 2001US-0339266P.  
 PR 24-OCT-2001; 2001US-0343629P.  
 PR 29-OCT-2001; 2001US-0349575P.  
 PR 01-NOV-2001; 2001US-0346357P.  
 PR 17-APR-2002; 2002US-0373260P.  
 PR 19-APR-2002; 2002US-0373815P.  
 PR 19-APR-2002; 2002US-0373817P.  
 PR 19-APR-2002; 2002US-0373826P.  
 PR 19-APR-2002; 2002US-0373884P.  
 PR 22-APR-2002; 2002US-0374977P.  
 PR 16-MAY-2002; 2002US-0381037P.  
 PR 16-MAY-2002; 2002US-0381038P.  
 PR 16-MAY-2002; 2002US-0381042P.  
 PR 17-MAY-2002; 2002US-0381642P.  
 PR 28-MAY-2002; 2002US-0383656P.  
 PR 29-MAY-2002; 2002US-0383831P.  
 PR 25-JUN-2002; 2002US-0391335P.  
 PR 01-OCT-2002; 2002US-00262511.  
 XX  
 XX (CURA-) CURAGEN CORP.  
 XX  
 XX Smithson G, Millet I, Peyman JA, Kekuda R, Ju J, Li L, Guo X;  
 PI Patturajan M, Szytek KA, Edinger SR, Ellerman K, Malyankar UM;  
 PI Ott T, Gorman L, Zerhusen BD, Anderson DM, Zhong M, Catterton E;  
 PI Ji W, Miller CE, Rastelli L, Stone DU, Pena CBA, Shenoy SG;

PI Shimkets RA, Rothenberg ME, Leach MD, Agee ML, Berghs C, Dipippo VA;  
 PI Eisen AJ, Gangolli EA, Rieger DK, Spaderna SK;  
 DR WPI; 2003-381626/36.  
 DR N-PSDB; ADA05731.  
 XX  
 PT New NOVX polypeptides and nucleic acids, useful for diagnosing,  
 PT preventing or treating NOVX-associated disorders, e.g. diabetes, obesity,  
 PT cancer or dyslipidemia, and in chromosome mapping, tissue typing or  
 PT pharmacogenomics.  
 XX  
 PS Claim 1; Page 169-170; 586pp; English.  
 XX  
 CC The present invention describes NOVX proteins, where X can be 1 to 55  
 CC (e.g. NOV1). Also described: (1) a composition comprising a polypeptide  
 CC described above and a carrier; (2) a kit comprising, in one or more  
 CC containers, the composition described above; (3) an isolated nucleic acid  
 CC molecule which encodes a NOVX protein of the invention; (4) a vector  
 CC comprising the nucleic acid molecule described above; (5) a cell  
 CC comprising the above vector; (6) an antibody that immunospecifically  
 CC binds to the polypeptide described above; (7) methods for determining the  
 CC presence or amount of the above polypeptide or nucleic acid molecule in a  
 CC sample; (8) methods for determining the presence of or predisposition to a  
 CC disease associated with altered levels of expression of the above  
 CC polypeptide or nucleic acid molecule in a first mammalian subject; (9) a  
 CC method of identifying an agent that binds to the polypeptide described  
 CC above; (10) a method for identifying a potential therapeutic agent for  
 CC use in treating a pathology that is related to an aberrant expression or  
 CC aberrant physiological interactions of the polypeptide; (11) a method of  
 CC screening for a modulator of activity or of latency or predisposition to  
 CC a pathology associated with the polypeptide; (12) a method for modulating  
 CC the activity of the polypeptide described above; (13) methods of treating  
 CC or preventing a pathology associated with the above polypeptide in a  
 CC mammal; and (14) a method for producing the above polypeptide. NOVX  
 CC sequences have antidiabetic, anorectic, antibacterial, virucide,  
 CC immunomodulator, cytostatic, nootropic, neuroprotective, antiparkinsonian  
 CC and antilipemic activities, and can be used in gene therapy. The  
 CC polypeptide is useful in manufacturing a medicament for treating a  
 CC syndrome associated with a human disease. The polypeptide or the nucleic  
 CC acid molecule may be used to diagnose, treat or prevent metabolic  
 CC disorders such as diabetes or obesity, infections, cachexia, cancer,  
 CC neurodegenerative disorders such as Alzheimer's disease or Parkinson's  
 CC disease, immune disorders, haematopoietic disorders and various  
 CC dyslipidaemias. The nucleic acids can also be used as hybridisation  
 CC probes, in chromosome mapping, tissue typing, preventive medicine and  
 CC pharmacogenomics. The present sequence represents a human NOVX from the  
 CC present invention.  
 XX  
 SQ Sequence 250 AA;  
 Query Match 100.0%; Score 47; DB 6; Length 250;  
 Best Local Similarity 100.0%; Pred. No. 2.5;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLVNERWVL 9  
 DB 55 VLVNERWVL 63

RESULT 10  
 ADA05734  
 ID ADA05734 standard; protein; 252 AA.  
 AC ADA05734;  
 XX  
 XX 06-NOV-2003 (first entry)  
 XX  
 DE Human NOV18b protein SEQ ID NO:94.  
 XX  
 KW human; NOVX; antidiabetic; anorectic; antibacterial; virucide;  
 KW immunomodulator; cytostatic; nootropic; neuroprotective;  
 KW antiparkinsonian; antilipemic; gene therapy; human disease;  
 KW metabolic disorder; diabetes; obesity; infection; cachexia; cancer;



neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;  
immune disorder; haematopoietic disorder; dyslipidaemia.

Homo sapiens.

WO2003029424-A2.

10-APR-2003.

02-OCT-2002; 2002WO-US031373.

02-OCT-2001; 2001US-0326483P.

05-OCT-2001; 2001US-0327435P.

05-OCT-2001; 2001US-0327449P.

09-OCT-2001; 2001US-0327917P.

09-OCT-2001; 2001US-0328029P.

09-OCT-2001; 2001US-0328044P.

09-OCT-2001; 2001US-0328056P.

12-OCT-2001; 2001US-0328849P.

15-OCT-2001; 2001US-0329414P.

17-OCT-2001; 2001US-0330142P.

18-OCT-2001; 2001US-0330309P.

22-OCT-2001; 2001US-0341058P.

24-OCT-2001; 2001US-0343629P.

01-NOV-2001; 2001US-0349575P.

29-OCT-2001; 2001US-0349575P.

17-APR-2002; 2002US-0373260P.

19-APR-2002; 2002US-0373815P.

19-APR-2002; 2002US-0373817P.

19-APR-2002; 2002US-0373826P.

19-APR-2002; 2002US-0373884P.

22-APR-2002; 2002US-0374977P.

16-MAY-2002; 2002US-0381037P.

16-MAY-2002; 2002US-0381038P.

16-MAY-2002; 2002US-0381042P.

28-MAY-2002; 2002US-0381642P.

28-MAY-2002; 2002US-0383656P.

29-MAY-2002; 2002US-0383831P.

25-JUN-2002; 2002US-0391335P.

01-OCT-2002; 2002US-00262511.

(CURA-) CURAGEN CORP.

Smithson G, Millet I, Peyman JA, Kekuda R, Ju J, Li L, Guo X;  
Patturajan M, Spytek KA, Edinger SR, Ellerman K, Malyankar UM;  
Ott T, Gorman L, Zerkhusen BD, Anderson DM, Zhong M, Carterton E;  
Ji W, Miller CE, Rastelli L, Stone DJ, Pena CE, Shenoy SG;  
Shinkets RA, Rotherberg ME, Leach MD, Agge ML, Berghs C, Dippio VA;  
Eisen AJ, Gangoli EA, Kieger DK, Spaderna SK;

WPI; 2003-381626/36.

N-PSDB; ADA05733.

New NOVX polypeptides and nucleic acids, useful for diagnosing,  
preventing or treating NOVX-associated disorders, e.g. diabetes, obesity,  
cancer or dyslipidemia, and in chromosome mapping, tissue typing or  
pharmacogenomics.

Claim 1; Page 170; 586pp; English.

The present invention describes NOVX proteins, where x can be 1 to 55  
(e.g. NOV1). Also described: (1) a composition comprising a polypeptide  
described above and a carrier; (2) a kit comprising, in one or more  
containers, the composition described above; (3) an isolated nucleic acid  
molecule which encodes a NOVX protein of the invention; (4) a vector  
comprising the nucleic acid molecule described above; (5) a cell  
comprising the above vector; (6) an antibody that immunospecifically  
binds to the polypeptide described above; (7) methods for determining the  
presence or amount of the above polypeptide or nucleic acid molecule in a  
sample; (8) methods for determining the presence of or predisposition to  
a disease associated with altered levels of expression of the above  
polypeptide or nucleic acid molecule in a first mammalian subject; (9) a

method of identifying an agent that binds to the polypeptide described  
above; (10) a method for identifying a potential therapeutic agent for  
use in treating a pathology that is related to an aberrant expression or  
aberrant physiological interactions of the polypeptide; (11) a method of  
screening for a modulator of activity or of latency or predisposition to  
a pathology associated with the polypeptide; (12) a method for modulating  
the activity of the polypeptide described above; (13) methods of treating  
or preventing a pathology associated with the above polypeptide in a  
mammal; and (14) a method for producing the above polypeptide. NOVX  
sequences have antidiabetic, anorectic, antibacterial, virucide,  
immunomodulator, cytostatic, neurotropic, neuroprotective, antiparkinsonian  
and antilipidemic activities, and can be used in gene therapy. The  
polypeptide is useful in manufacturing a medicament for treating a  
syndrome associated with a human disease. The polypeptide or the nucleic  
acid molecule may be used to diagnose, treat or prevent metabolic  
disorders such as diabetes or obesity, infections, cachexia, cancer,  
neurodegenerative disorders such as Alzheimer's disease or Parkinson's  
disease, immune disorders, haematopoietic disorders and various  
dyslipidaemias. The nucleic acids can also be used as hybridisation  
probes, in chromosome mapping, tissue typing, preventive medicine and  
pharmacogenomics. The present sequence represents a human NOVX from the  
present invention.

Sequence 252 AA;

Query Match 100.0%; Score 47; DB 6; Length 252;  
Best Local Similarity 100.0%; Pred. No. 2.5;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLVNERWVL 9  
Dd 57 VLVNERWVL 65

RESULT 11  
AAR67888  
ID AAR67888 standard; protein; 253 AA.

XX AAR67888;  
AC AAR67888;  
XX 25-MAR-2003 (revised)  
DT 09-AUG-1995 (first entry)

XX Human stratum corneum chymotrophic recombinant enzyme (SCCE).  
XX Stratum corneum chymotryptic enzyme; skin disorder; acne; psoriasis;  
KW callosities; keratosis pilaris; ichthyoses; eczema.  
XX Homo sapiens.  
XX W09500651-A1.  
PN 05-JAN-1995.  
XX 20-JUN-1994; 94WO-1B000166.  
PF 18-JUN-1993; 93DK-00000725.  
PR (SYMB-) SYMBICOM AB.  
XX Egelrud T, Hansson L;  
PI WPI; 1995-052088/07.  
XX N-PSDB; AAQ81203.  
PT Nucleotide sequences encoding stratum corneum chymotryptic enzyme - and  
PT related vectors, transformed cells and polypeptides, useful for treating  
PT skin disorders, e.g. acne or psoriasis, and for identification of  
XX specific inhibitors.  
XX Disclosure; Page 97; 137pp; English.  
XX The enzyme encoded by this sequence is used in pharmaceutical, cosmetic

C and skin care products, especially to treat and prevent acne, xeroderma,  
 C or other hyperkeratotic conditions (e.g. callosities or keratosis  
 C pilaris), ichthyoses, psoriasis, eczema, etc. It is produced  
 C recombinantly following mammal, insect, plant, or microorganism  
 C transformation with plasmid pS507. (Updated on 25-MAR-2003 to correct PN  
 C field.)  
 X  
 X

Q Sequence 253 AA;

Query Match 100.0%; Score 47; DB 2; Length 253;  
 Best Local Similarity 100.0%; Pred. No. 2.5;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2Y 1 VLVNERWVL 9  
 |||||  
 58 VLVNERWVL 66

## RESULT 12

AAW05383  
 ID AAW05383 standard; protein; 253 AA.

AC AAW05383;

DT 31-DEC-1996 (first entry);

DE Human amyloid precursor protein protease.

KW Amyloid precursor protein protease; Alzheimer's disease; diagnosis;  
 KW therapy.

OS Homo sapiens.

PN WO9631122-A1.

PD 10-OCT-1996.

PP 02-APR-1996; 96WO-US004294.

PR 04-APR-1995; 95US-00416257.

PA (ELIL) LILLY & CO ELI.

PI Dixon BP, Johnstone EM, Little SP;

DR WPI; 1996-464694/46.

DR N-PSDB; AAT39783.

PT New isolated human amyloid precursor protein protease - used to develop  
 PT prods; for the treatment or diagnosis of associated conditions, esp.  
 PT Alzheimer's disease.

PS Claim 1; Page 44-45; 55pp; English.

CC Human amyloid precursor protein protease (AAW05383) is involved in the  
 CC processing or clearance of amyloid precursor protein to form beta-amyloid  
 CC peptide. Its amino acid sequence was deduced from a cDNA clone (AAT39783)  
 CC obd. from a human lung library. Recombinant protease can be produced in  
 CC transformed or transfected prokaryotic (partic. E. coli) or eukaryotic  
 CC (partic. AV-120 host cells). It is used to develop products for the design  
 CC and testing of cpds. useful for treating or preventing conditions  
 CC associated with beta-amyloid peptide, esp. Alzheimer's disease

SQ Sequence 253 AA;

Query Match 100.0%; Score 47; DB 2; Length 253;  
 Best Local Similarity 100.0%; Pred. No. 2.5;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLVNERWVL 9  
 |||||  
 58 VLVNERWVL 66

## RESULT 13

ABB84421

ID ABB84421 standard; peptide; 253 AA.

XX ABB84421;

XX 08-NOV-2002 (first entry)

XX Human SCCE protein N-terminal fragment SEQ ID 48.

XX SCCE; human; stratum corneum chymotryptic enzyme; kallikrein 7;

XX serine protease; transgenic mammal; skin; skin disease; skin cancer;

XX hyperkeratosis; acanthosis; epidermal inflammation; dermal inflammation;

XX pruritus; atopic dermatitis; eczema; acne; itch; KIK7; enzyme.

OS Homo sapiens.

PN WO200262135-A2.

PD 15-AUG-2002.

PP 08-FEB-2002; 2002WO-IB001300.

PR 09-FEB-2001; 2001CA-02332655.

PR 09-FEB-2001; 2001DK-00000218.

PA (EGEL/) EGELRUD T.

PA (HANS/) HANSSON L.

PI Egelrud T, Hansson L;

DR WPI; 2002-643380/69.

PT Transgenic mammal or its embryo useful as model for human disease, has  
 PT heterologous nucleotide sequence coding for stratum corneum chymotryptic  
 PT enzyme operably linked to promoter that drives its expression in skin.

XX Example 6; Page 37; 74pp; English.

CC This invention describes a novel non-human transgenic mammal or mammalian  
 CC embryo having integrated within its genome, a heterologous nucleotide  
 CC sequence comprising at least a significant part of a nucleotide sequence  
 CC coding for a stratum corneum chymotryptic enzyme (SCCE) or its variant,  
 CC operably linked to a promoter that drives expression of heterologous scce  
 CC or its variant in skin. The product of the invention is useful as a model  
 CC for the study of disease with the aim of improving treatment, to relieve  
 CC or ameliorate a pathogenic condition, for development or testing of a  
 CC cosmetic or a pharmaceutical formulation, and for the development of a  
 CC diagnostic method. It can also be used as a model for a skin disease or  
 CC skin cancer. The invention is also useful for screening or identifying a  
 CC compound or composition effective for the prevention or treatment of an  
 CC abnormal or unwanted phenotype, and for screening or identifying a  
 CC compound or composition effective for the prevention or treatment of  
 CC inflammatory skin diseases selected from diseases consisting of epidermal  
 CC hyperkeratosis, acanthosis, epidermal inflammation, dermal inflammation,  
 CC pruritus, atopic dermatitis, eczema, acne and inherited skin diseases  
 CC with epidermal hyperkeratosis. The mammal of the invention is also useful  
 CC as a model for further studies of itch mechanisms and the testing of  
 CC potential compounds and compositions for relieve of various skin diseases  
 CC where itch is a component. This sequence represents the N-terminal  
 CC fragment of the human stratum corneum chymotryptic enzyme, SCCE  
 CC synonymous with human kallikrein 7 (KIK7), used in the development of the  
 CC transgenic mammals described in the invention

SQ Sequence 253 AA;

Query Match 100.0%; Score 47; DB 5; Length 253;  
 Best Local Similarity 100.0%; Pred. No. 2.5;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLVNERWVL 9  
 |||||

58 VLVNERWVL 66

RESULT 14

AB84406  
 ID ABB84406 standard; protein; 253 AA.

XX ABB84406;

XX 08-NOV-2002 (first entry)

XX Human SCCE protein.

XX SCCE; human; stratum corneum chymotryptic enzyme; kallikrein 7;  
 KW serine protease; transgenic mammal; skin; skin disease; skin cancer;  
 KW hyperkeratosis; acanthosis; epidermal inflammation; dermal inflammation;  
 KW pruritus; atopic dermatitis; eczema; acne; itch; KLK7; enzyme.

XX Homo sapiens.

XX WO200262135-A2.

XX 15-AUG-2002.

XX 08-FEB-2002; 2002WO-1B001300.

XX 09-FEB-2001; 2001CA-02332655.

XX 09-FEB-2001; 2001DR-00000218.

XX (EGEL/) EGELRUD T.

XX (HANS/) HANSSON L.

XX Egelrud T, Hansson L;

XX WPI; 2002-643380/69.

XX N-PSDB; ABQ76226.

XX Transgenic mammal or its embryo useful as model for human disease, has  
 PT heterologous nucleotide sequence coding for stratum corneum chymotryptic  
 PT enzyme operably linked to promoter that drives its expression in skin.

XX Claim 10; Page 58-59; 74pp; English.

XX This invention describes a novel non-human transgenic mammal or mammalian  
 CC embryo having integrated within its genome, a heterologous nucleotide  
 CC sequence comprising at least a significant part of a nucleotide sequence  
 CC coding for a stratum corneum chymotryptic enzyme (SCCE) or its variant,  
 CC operably linked to a promoter that drives expression of heterologous scce  
 CC or its variant in skin. The product of the invention is useful as a model  
 CC for the study of disease with the aim of improving treatment, to relieve  
 CC or ameliorate a pathogenic condition, for development or testing of a  
 CC cosmetic or a pharmaceutical formulation, and for the development of a  
 CC diagnostic method. It can also be used as a model for a skin disease or  
 CC skin cancer. The invention is also useful for screening or identifying a  
 CC compound or composition effective for the prevention or treatment of an  
 CC abnormal or unwanted phenotype, and for screening or identifying a  
 CC compound or composition effective for the prevention or treatment of  
 CC inflammatory skin diseases selected from diseases consisting of epidermal  
 CC hyperkeratosis, acanthosis, epidermal inflammation, dermal inflammation,  
 CC pruritus, atopic dermatitis, eczema, acne and inherited skin diseases  
 CC with epidermal hyperkeratosis. The mammal of the invention is also useful  
 CC as a model for further studies of its mechanisms and the testing of  
 CC potential compounds and compositions for relief of various skin diseases  
 CC where itch is a component. This sequence represents the human stratum  
 CC corneum chymotryptic enzyme, SCCE which is a serine protease synonymous  
 CC with human kallikrein 7 (KLK7) and is used in the development of the  
 CC transgenic mammals described in the invention

XX Sequence 253 AA;

Query Match 100.0%; Score 47; DB 5; Length 253;

Best Local Similarity 100.0%; Pred. No. 2.5;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLVNERWVL 9

DB 58 VLVNERWVL 66

RESULT 15

AAU82740  
 ID AAU82740 standard; protein; 253 AA.

XX AAU82740;

XX 23-APR-2002 (first entry)

XX Amino acid sequence of novel human protease #39.

XX Human; protease; cancer; immune-related disorder; cardiovascular disease;  
 KW neuronal-associated disease; metabolic disorder; inflammatory disorder;  
 KW nervous system disorder; sexual dysfunction; pain; mood disorder;  
 KW hypertension; psychotic disorder; neurological disorder; dyskinesia;  
 KW viral infection; human immunodeficiency virus; HIV; non-viral infection;  
 KW ocular disease; cytostatic; enzyme.

XX Homo sapiens.

XX WO200200860-A2.

XX 03-JAN-2002.

XX 26-JUN-2001; 2001WO-US020171.

XX 26-JUN-2000; 2000US-0214047P.

XX (SUGE-) SUGEN INC.

XX Plowman G, Whyte D, Sudarsanam S, Manning G, Caenepeel S;

XX Charyczak G;

XX WPI; 2002-139913/18.

XX N-PSDB; ABK31782.

XX Nucleic acids encoding novel human proteases, useful for useful for  
 PT treating diseases and disorders such as cancers, immune-related diseases  
 PT and disorders, cardiovascular disease (e.g. restenosis) and inflammatory  
 PT disorders.

XX Claim 6; Fig 2N; 313pp; English.

XX The present invention relates to the isolation of novel human proteases,  
 CC and the nucleic acids encoding them. The sequences of the invention are  
 CC useful for treating diseases and disorders such as cancers (e.g. breast,  
 CC colon, lung), immune-related diseases and disorders (e.g. inflammatory  
 CC diseases and asthma), cardiovascular diseases (e.g. restenosis and  
 CC coronary thrombosis), brain or neuronal-associated diseases, metabolic  
 CC disorders (e.g. diabetes, obesity), inflammatory disorders (e.g.  
 CC rheumatoid arthritis and psoriasis), central or peripheral nervous system  
 CC diseases, migraines, pain, sexual dysfunction, mood disorders, attention  
 CC disorders, cognitive disorders, hypotension, hypertension, psychotic  
 CC disorders, neurological disorders (e.g. Alzheimer's disease, Parkinson's  
 CC disease) and dyskinesias. The nucleic acids and polypeptides are also  
 CC useful for treating viral infections caused by human immunodeficiency  
 CC virus (HIV), and non-viral infections such as ocular disease (e.g.  
 CC glaucoma) and macular degeneration. AAU82702-AAU82760 represent the novel  
 CC human proteases of the invention

XX Sequence 253 AA;

Query Match 100.0%; Score 47; DB 5; Length 253;

Best Local Similarity 100.0%; Pred. No. 2.5;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLVNERWVL 9

Wed Mar 3 07:49:11 2004

> 58 VLVNERWVL 66

search completed: March 1, 2004, 17:28:51  
db time : 47.5556 secs



APPLICANT: O'Brien, Timothy J.  
 TITLE OF INVENTION: Compositions and Methods for the Early Diagnosis of  
 TITLE OF INVENTION: Ovarian Cancer  
 FILE REFERENCE: D6223CIP/C/Div  
 CURRENT APPLICATION NUMBER: US/09/905,083  
 PRIOR FILING DATE: 2001-07-13  
 PRIOR APPLICATION NUMBER: US 09/502,600  
 PRIOR FILING DATE: 2000-02-11  
 NUMBER OF SEQ ID NOS: 136  
 SEQ ID NO 34  
 LENGTH: 9  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 FEATURE:  
 NAME/KEY: CHAIN  
 OTHER INFORMATION: Residues 58-66 of the SCE protein  
 09-905-083-34

Query Match 100.0%; Score 47; DB 9; Length 9;  
 Best Local Similarity 100.0%; Pred. No. 7.1e+05;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 VLVNERWVL 9  
 |||||  
 1 VLVNERWVL 9

RESULT 3  
 Sequence 98, Application US/09888615  
 Patent No. US20020064856A1  
 GENERAL INFORMATION:  
 APPLICANT: PLOWMAN, GREGORY  
 APPLICANT: WHYTE, DAVID  
 APPLICANT: CAENEPEEL, SEAN  
 APPLICANT: CHARYDCZAK, GLEN  
 APPLICANT: MANNING, GERARD  
 APPLICANT: SUDARSANAM, SUCHA  
 TITLE OF INVENTION: NOVEL PROTEASES  
 FILE REFERENCE: 038602/1214  
 CURRENT APPLICATION NUMBER: US/09/888,615  
 CURRENT FILING DATE: 2001-06-26  
 PRIOR APPLICATION NUMBER: 60/214,047  
 PRIOR FILING DATE: 2000-06-26  
 NUMBER OF SEQ ID NOS: 150  
 SOFTWARE: Patentin Ver. 2.1  
 SEQ ID NO 98  
 LENGTH: 253  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 US-09-888-615-98

Query Match 100.0%; Score 47; DB 9; Length 253;  
 Best Local Similarity 100.0%; Pred. No. 2.9;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLVNERWVL 9  
 |||||  
 Db 58 VLVNERWVL 66

RESULT 4  
 US-09-764-762-3  
 Sequence 5, Application US/09764762  
 Patent No. US20020069341A1  
 GENERAL INFORMATION:  
 APPLICANT: Hillman, Jennifer L.  
 TITLE OF INVENTION: NOVEL KALLIKREIN  
 NUMBER OF SEQUENCES: 5  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Incyte Pharmaceuticals, Inc.  
 STREET: 3174 Porter Drive

CITY: Palo Alto  
 STATE: CA  
 COUNTRY: USA  
 ZIP: 94304  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette  
 COMPUTER: IBM Compatible  
 OPERATING SYSTEM: DOS  
 SOFTWARE: FASTSEQ for Windows Version 2.0  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/764,762  
 FILING DATE: 16-Jan-2001  
 CLASSIFICATION: <Unknown>  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 09/210,084  
 FILING DATE: <Unknown>  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Billings, Lucy J.  
 REGISTRATION NUMBER: 36,749  
 REFERENCE/DOCKET NUMBER: PP-0252 US  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 415-855-0555  
 TELEFAX: 415-845-4166  
 INFORMATION FOR SEQ ID NO: 3:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 253 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 IMMEDIATE SOURCE:  
 LIBRARY: GenBank  
 CLONE: 532504  
 SEQUENCE DESCRIPTION: SEQ ID NO: 3:  
 US-09-764-762-3

Query Match 100.0%; Score 47; DB 9; Length 253;  
 Best Local Similarity 100.0%; Pred. No. 2.9;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLVNERWVL 9  
 |||||  
 Db 58 VLVNERWVL 66

RESULT 5  
 US-10-264-283-90  
 Sequence 90, Application US/10264283  
 Publication No. US2003014494A1  
 GENERAL INFORMATION:  
 APPLICANT: Algate, Paul A.  
 APPLICANT: Mannion, Jane  
 TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
 FILE REFERENCE: 210121.590  
 CURRENT APPLICATION NUMBER: US/10/264,283  
 CURRENT FILING DATE: 2002-10-02  
 NUMBER OF SEQ ID NOS: 111  
 SOFTWARE: Corixa Invention Disclosure Database  
 SEQ ID NO 90  
 LENGTH: 253  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 US-10-264-283-90

Query Match 100.0%; Score 47; DB 14; Length 253;  
 Best Local Similarity 100.0%; Pred. No. 2.9;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLVNERWVL 9  
 |||||  
 Db 58 VLVNERWVL 66

LT 5  
0-295-027-498  
Sequence 498, Application US/10295027  
Publication No. US2003023250A1  
GENERAL INFORMATION:

APPLICANT: Afar, Daniel  
APPLICANT: Aziz, Natasha  
APPLICANT: Ginsberg, Wendy M.  
APPLICANT: Gish, Kurt C.  
APPLICANT: Glynn, Richard  
APPLICANT: Hevezi, Peter A.  
APPLICANT: Mack, David H.  
APPLICANT: Murray, Richard  
APPLICANT: Watson, Susan R.  
APPLICANT: Eos Biotechnology, Inc.

TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and  
TITLE OF INVENTION: Methods of Screening for Modulators of Cancer

FILE REFERENCE: 018501-012500US  
CURRENT APPLICATION NUMBER: US/10/295,027

CURRENT FILING DATE: 2002-11-13  
PRIOR FILING DATE: 2000-09-15  
PRIOR FILING DATE: 2000-09-15  
PRIOR FILING DATE: 2001-11-13  
PRIOR FILING DATE: 2001-11-13  
PRIOR FILING DATE: 2001-11-15  
PRIOR FILING DATE: 2001-11-21  
PRIOR FILING DATE: 2001-11-29  
PRIOR FILING DATE: 2001-12-14  
PRIOR FILING DATE: 2002-01-08  
PRIOR FILING DATE: 2002-01-10  
PRIOR FILING DATE: 2002-02-08  
PRIOR FILING DATE: 2002-02-13  
PRIOR FILING DATE: 2002-02-13  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 1386  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 498  
LENGTH: 253  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-295-027-498

Query Match 100.0%; Score 47; DB 15; Length 253;  
Best Local Similarity 100.0%; Pred. No. 2.9; 0; Indels 0; Gaps 0;  
Matches 9; Conservative 0; Mismatches 0;

Qy 1 VLVNERWVL 9  
Db 58 VLVNERWVL 66

RESULT 7  
US-10-173-999-48  
Sequence 48, Application US/10173999  
Publication No. US20040005563A1

GENERAL INFORMATION:  
APPLICANT: Mack, David H.  
APPLICANT: Gish, Kurt C.  
APPLICANT: Eos Biotechnology, Inc.

TITLE OF INVENTION: Methods of Diagnosis of Ovarian Cancer, Compositions  
TITLE OF INVENTION: Methods of Screening for Modulators of Ovarian  
TITLE OF INVENTION: Cancer  
FILE REFERENCE: 018501-002420US  
CURRENT APPLICATION NUMBER: US/10/173,999  
CURRENT FILING DATE: 2002-06-17  
PRIOR APPLICATION NUMBER: US 60/299,234

PRIOR FILING DATE: 2001-06-18  
PRIOR APPLICATION NUMBER: US 60/315,287  
PRIOR FILING DATE: 2001-08-27  
PRIOR APPLICATION NUMBER: US 60/350,666  
PRIOR FILING DATE: 2001-11-13  
PRIOR APPLICATION NUMBER: US 60/372,246  
PRIOR FILING DATE: 2001-04-12  
NUMBER OF SEQ ID NOS: 163  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 48  
LENGTH: 253  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-173-999-48

Query Match 100.0%; Score 47; DB 15; Length 253;  
Best Local Similarity 100.0%; Pred. No. 2.9; 0; Indels 0; Gaps 0;  
Matches 9; Conservative 0; Mismatches 0;

Qy 1 VLVNERWVL 9  
Db 58 VLVNERWVL 66

RESULT 8  
US-09-918-243-40

Sequence 40, Application US/09918243  
Patent No. US20020142317A1  
GENERAL INFORMATION:  
APPLICANT: O'Brien, Timothy J.  
APPLICANT: Cannon, Martin J.  
APPLICANT: Santin, Alessandro  
TITLE OF INVENTION: Methods for the early diagnosis of ovarian cancer  
FILE REFERENCE: D6223CIP/C/D/CIP/09/918,243  
CURRENT APPLICATION NUMBER: US/09/918,243  
CURRENT FILING DATE: 2001-07-30  
PRIOR APPLICATION NUMBER: US  
PRIOR FILING DATE: 2001-07-13  
NUMBER OF SEQ ID NOS: 136  
SEQ ID NO 40  
LENGTH: 9  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CHAIN  
OTHER INFORMATION: Residues 57-65 of the SCCE protein  
US-09-918-243-40

Query Match 91.5%; Score 43; DB 9; Length 9;  
Best Local Similarity 100.0%; Pred. No. 7.1e+05; 0; Indels 0; Gaps 0;  
Matches 8; Conservative 0; Mismatches 0;

Qy 1 VLVNERWV 8  
Db 2 VLVNERWV 9

RESULT 9  
US-09-918-243-50

Sequence 50, Application US/09918243  
Patent No. US20020142317A1  
GENERAL INFORMATION:  
APPLICANT: O'Brien, Timothy J.  
APPLICANT: Cannon, Martin J.  
APPLICANT: Santin, Alessandro  
TITLE OF INVENTION: Methods for the early diagnosis of ovarian cancer  
FILE REFERENCE: D6223CIP/C/D/CIP/09/918,243  
CURRENT APPLICATION NUMBER: US/09/918,243  
CURRENT FILING DATE: 2001-07-30  
PRIOR APPLICATION NUMBER: US  
PRIOR FILING DATE: 2001-07-13  
NUMBER OF SEQ ID NOS: 136  
SEQ ID NO 50

LENGTH: 9  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 FEATURE:  
 NAME/KEY: CHAIN  
 OTHER INFORMATION: Residues 59-67 of the SCCE protein  
 09-918-243-50

Query Match 91.5%; Score 43; DB 9; Length 9;  
 Best Local Similarity 100.0%; Pred. No. 7.1e+05;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0;  
 Gaps 0;

2 LVNERWVL 9  
 1 LVNERWVL 8

## SULT 10

--09-905-083-40  
 Sequence 40, Application US/09905083  
 Patent No. US20020146708A1  
 GENERAL INFORMATION:  
 APPLICANT: O'Brien, Timothy J.  
 TITLE OF INVENTION: Compositions and Methods for the Early Diagnosis of  
 TITLE OF INVENTION: Ovarian Cancer  
 FILE REFERENCE: D6223CIP/C/Div  
 CURRENT APPLICATION NUMBER: US/09/905,083  
 PRIOR FILING DATE: 2001-07-13  
 PRIOR APPLICATION NUMBER: US 09/502,600  
 PRIOR FILING DATE: 2000-02-11  
 NUMBER OF SEQ ID NOS: 136  
 SEQ ID NO 40  
 LENGTH: 9  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 FEATURE:  
 NAME/KEY: CHAIN  
 OTHER INFORMATION: Residues 57-65 of the SCCE protein  
 US-09-905-083-40

Query Match 91.5%; Score 43; DB 9; Length 9;  
 Best Local Similarity 100.0%; Pred. No. 7.1e+05;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0;  
 Gaps 0;

1 VLVNERWV 8  
 2 VLVNERWV 9

## RESULT 11

US-09-905-083-50  
 Sequence 50, Application US/09905083  
 Patent No. US20020146708A1  
 GENERAL INFORMATION:  
 APPLICANT: O'Brien, Timothy J.  
 TITLE OF INVENTION: Compositions and Methods for the Early Diagnosis of  
 TITLE OF INVENTION: Ovarian Cancer  
 FILE REFERENCE: D6223CIP/C/Div  
 CURRENT APPLICATION NUMBER: US/09/905,083  
 PRIOR FILING DATE: 2001-07-13  
 PRIOR APPLICATION NUMBER: US 09/502,600  
 PRIOR FILING DATE: 2000-02-11  
 NUMBER OF SEQ ID NOS: 136  
 SEQ ID NO 50  
 LENGTH: 9  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 FEATURE:  
 NAME/KEY: CHAIN  
 OTHER INFORMATION: Residues 59-67 of the SCCE protein  
 US-09-905-083-50

Query Match 91.5%; Score 43; DB 9; Length 9;

Best Local Similarity 100.0%; Pred. No. 7.1e+05;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0;  
 Gaps 0;

QY 2 LVNERWVL 9  
 DB 1 LVNERWVL 8

## RESULT 12

US-10-202-676-6  
 Sequence 6, Application US/10202676  
 Publication No. US20030092620A1  
 GENERAL INFORMATION:  
 APPLICANT: Lucas, John  
 TITLE OF INVENTION: USE OF ADIPSIN/COMPLEMENT FACTOR D IN THE TREATMENT OF METABOLIC  
 TITLE OF INVENTION: RELATED DISORDERS  
 FILE REFERENCE: 121.US2.REG  
 CURRENT APPLICATION NUMBER: US/10/202,676  
 PRIOR FILING DATE: 2002-07-22  
 PRIOR APPLICATION NUMBER: 60/308,147  
 PRIOR FILING DATE: 2001-07-26  
 NUMBER OF SEQ ID NOS: 6  
 SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO 6  
 LENGTH: 228  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 FEATURE:  
 NAME/KEY: VARIANT  
 LOCATION: 27  
 OTHER INFORMATION: Polymorphic amino acid Ala or Gly  
 FEATURE:  
 NAME/KEY: VARIANT  
 LOCATION: 34  
 OTHER INFORMATION: Polymorphic amino acid Arg or Gln  
 FEATURE:  
 NAME/KEY: VARIANT  
 LOCATION: 125  
 OTHER INFORMATION: Polymorphic amino acid Val or Leu  
 US-10-202-676-6

Query Match 83.0%; Score 39; DB 14; Length 228;  
 Best Local Similarity 88.9%; Pred. No. 55;  
 Matches 8; Conservative 0; Mismatches 1; Indels 0;  
 Gaps 0;

QY 1 VLVNERWVL 9  
 DB 29 VLVNERWVL 37

## RESULT 13

US-10-045-367A-7  
 Sequence 7, Application US/10045367A  
 Publication No. US20020165152A1  
 GENERAL INFORMATION:  
 APPLICANT: Kapeller-Libermann, Rosana  
 TITLE OF INVENTION: 14089, A NOVEL HUMAN TRYPSIN SERINE  
 TITLE OF INVENTION: PROTEASE AND USES THEREOF  
 FILE REFERENCE: 10448-109001  
 CURRENT APPLICATION NUMBER: US/10/045,367A  
 PRIOR FILING DATE: 2002-04-15  
 PRIOR APPLICATION NUMBER: 60/246,561  
 PRIOR FILING DATE: 2000-11-07  
 NUMBER OF SEQ ID NOS: 7  
 SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO 7  
 LENGTH: 81  
 TYPE: PRT  
 ORGANISM: Artificial Sequence  
 FEATURE:  
 OTHER INFORMATION: Consensus sequence  
 US-10-045-367A-7



ry Match 80.9%; Score 38; DB 13; Length 81;  
t Local Similarity 75.0%; Pred. No. 32;  
ches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

2 LVNWRWVL 9  
|:|:|:|:|  
5 LINEQWVL 12

LT 14

0-170-789-57  
quence S7, Application US/10170789  
lication No. US20030180930A1  
GENERAL INFORMATION:  
APPLICANT: Rachel E. Meyers  
APPLICANT: Olandt, Peter J.  
APPLICANT: Kapeller-Libermann, Rosana  
APPLICANT: Curtis, Rory A. J.  
APPLICANT: Williamson, Mark  
APPLICANT: Welch, Nadine  
TITLE OF INVENTION: NOVEL HUMAN PROTEIN KINASE, PHOSPHATASE,  
TITLE OF INVENTION: AND PROTEASE FAMILY MEMBERS AND USES THEREOF  
FILE REFERENCE: 10448-191001.  
CURRENT APPLICATION NUMBER: US/10170,789

CURRENT FILING DATE: 2002-06-13  
PRIOR APPLICATION NUMBER: US 09/797,039  
PRIOR FILING DATE: 2001-02-28  
PRIOR APPLICATION NUMBER: PCT/US01/06525  
PRIOR FILING DATE: 2001-02-28  
PRIOR APPLICATION NUMBER: US 60/186,061  
PRIOR FILING DATE: 2000-02-29  
PRIOR APPLICATION NUMBER: US 09/882,166  
PRIOR FILING DATE: 2001-06-15  
PRIOR APPLICATION NUMBER: PCT/US01/19269  
PRIOR FILING DATE: 2001-06-15  
PRIOR APPLICATION NUMBER: US 60/212,078  
PRIOR FILING DATE: 2000-06-15  
PRIOR APPLICATION NUMBER: US 09/934,406  
PRIOR FILING DATE: 2001-08-21  
PRIOR APPLICATION NUMBER: PCT/US01/26052  
PRIOR FILING DATE: 2001-08-21  
PRIOR APPLICATION NUMBER: US 60/226,740  
PRIOR FILING DATE: 2000-08-21  
PRIOR APPLICATION NUMBER: US 09/861,801  
PRIOR FILING DATE: 2001-05-21  
PRIOR APPLICATION NUMBER: PCT/US01/16549  
PRIOR FILING DATE: 2001-05-21  
PRIOR APPLICATION NUMBER: US 60/205,508  
PRIOR FILING DATE: 2000-05-19  
PRIOR APPLICATION NUMBER: US 09/801,267  
PRIOR FILING DATE: 2001-03-06  
PRIOR APPLICATION NUMBER: PCT/US01/07138  
PRIOR FILING DATE: 2001-03-05  
PRIOR APPLICATION NUMBER: US 60/187,454  
PRIOR FILING DATE: 2000-03-07  
PRIOR APPLICATION NUMBER: US 09/829,671  
PRIOR FILING DATE: 2001-04-10  
PRIOR APPLICATION NUMBER: PCT/US01/40483  
PRIOR FILING DATE: 2001-04-11  
PRIOR APPLICATION NUMBER: US 60/197,508  
PRIOR FILING DATE: 2000-04-18  
PRIOR APPLICATION NUMBER: US 09/961,721  
PRIOR FILING DATE: 2001-09-24  
PRIOR APPLICATION NUMBER: PCT/US01/29904  
PRIOR FILING DATE: 2001-09-24  
PRIOR APPLICATION NUMBER: US 60/235,023  
PRIOR FILING DATE: 2000-09-25  
PRIOR APPLICATION NUMBER: US 10/045,367  
PRIOR FILING DATE: 2001-11-07  
PRIOR APPLICATION NUMBER: US 60/246,561  
PRIOR FILING DATE: 2000-11-07  
PRIOR APPLICATION NUMBER: US 09/801,275

PRIOR FILING DATE: 2001-03-06  
PRIOR APPLICATION NUMBER: PCT/US01/07074  
PRIOR FILING DATE: 2001-03-05  
PRIOR APPLICATION NUMBER: US 60/187,420  
PRIOR FILING DATE: 2000-03-07  
NUMBER OF SEQ ID NOS: 63  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 57  
LENGTH: 81  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: consensus sequence  
US-10-170-789-57

Query Match 80.9%; Score 38; DB 14; Length 81;  
Best Local Similarity 75.0%; Pred. No. 32;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LVNWRWVL 9  
|:|:|:|:|  
DB 5 LINEQWVL 12

RESULT 15

US-09-796-294-8  
Sequence 8, Application US/09796294  
Patent No. US20020037581A1  
GENERAL INFORMATION:  
APPLICANT: O'Brien, Timothy J.  
APPLICANT: Underwood, Lowell J.  
TITLE OF INVENTION: Extracellular Serine Protease  
FILE REFERENCE: D6020C1P3  
CURRENT APPLICATION NUMBER: US/09/796,294  
CURRENT FILING DATE: 2001-02-28  
PRIOR APPLICATION NUMBER: US 09/618,259  
PRIOR FILING DATE: 2000-07-18  
NUMBER OF SEQ ID NOS: 72  
SEQ ID NO 8  
LENGTH: 260  
TYPE: PRT  
ORGANISM: Mus sp.  
FEATURE:  
OTHER INFORMATION: Amino acid sequence of mouse neuropsin homologous  
to TADG-14; accession no. D30785  
US-09-796-294-8

Query Match 80.9%; Score 38; DB 9; Length 260;  
Best Local Similarity 77.8%; Pred. No. 91;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VLVNWRWVL 9  
|:|:|:|:|  
DB 61 VLVGRWVL 69

Search completed: March 1, 2004, 18:08:50  
Job time : 24.1111 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

rotein - protein search, using sw model

on: March 1, 2004, 17:21:01 ; Search time 10.2222 Seconds  
(without alignments)  
84.690 Million cell updates/sec

le: US-09-905-083-34

fect score: 47

uence: 1 VLVNERWVL 9

ring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

arched: 283366 seqs, 96191526 residues

al number of hits satisfying chosen parameters: 283366

imum DB seq length: 0

ximum DB seq length: 2000000000

st-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

itabase :

1: Pir1.\*

2: Pir2.\*

3: Pir3.\*

4: Pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

result No.	Score	Query Match	Length	ID	Description
1	47	100.0	253	2 A53968	serine proteinase
2	39	83.0	246	1 DBHU	complement factor
3	38	80.9	232	1 KOPG	tissue kallikrein
4	38	80.9	260	2 156559	neuropsin - mouse
5	38	80.9	417	2 T25594	hypothetical prote
6	37	78.7	250	2 T01279	trypsin (EC 3.4.21
7	37	78.7	250	2 S31384	trypsin (EC 3.4.21
8	37	78.7	258	1 S44184	chymotrypsin (EC 3
9	37	78.7	259	1 S49129	chymotrypsin (EC 3
10	37	78.7	1421	2 T05892	hypothetical prote
11	36	76.6	72	2 F82825	hypothetical prote
12	36	76.6	258	2 A45161	serine proteinase
13	36	76.6	426	2 T28563	hypothetical prote
14	36	76.6	426	2 C36850	A21R protein - var
15	36	76.6	426	2 C72166	A22R protein - var
16	36	76.6	426	2 T37408	probable 49.1K pro
17	36	76.6	426	2 D42519	A20R protein - vac
18	36	76.6	516	2 T47075	hypothetical prote
19	35	74.5	178	2 T25570	hypothetical prote
20	35	74.5	231	2 S31178	trypsin (EC 3.4.21
21	35	74.5	242	2 S31775	trypsin (EC 3.4.21
22	35	74.5	242	2 S49489	trypsin (EC 3.4.21
23	35	74.5	242	2 S31776	trypsin (EC 3.4.21
24	35	74.5	242	2 S31775	trypsin (EC 3.4.21
25	35	74.5	257	1 UC2479	venomatin B (EC 3.4
26	35	74.5	258	2 T36945	haptoglobin Hp - c
27	35	74.5	258	2 T36947	haptoglobin Hpp - c
28	35	74.5	258	2 G02959	haptoglobin - rhes
29	35	74.5	274	2 T47078	coagulation factor

ribosomal protein  
hypothetical prote  
haptoglobin precu  
haptoglobin - chim  
haptoglobin - chim  
haptoglobin - chim  
haptoglobin - precu  
haptoglobin - precu  
haptoglobin - precu  
haptoglobin - relat  
haptoglobin - chim  
haptoglobin - chim  
haptoglobin - chim  
branched-chain-ami  
coagulation factor  
acrosin (EC 3.4.21  
hypothetical prote  
unknown protein li  
probable membrane

## ALIGNMENTS

### RESULT 1

A53968  
serine proteinase SCCE precursor - human  
N/Alternate names: stratum corneum chymotryptic enzyme  
C/Species: Homo sapiens (man)  
C/Date: 07-Jul-1995 #sequence\_revision 07-Jul-1995 #text\_change 22-Jun-1999  
R/Accession: A53968  
R/Hansson, L.; Stroemqvist, M.; Baeckman, A.; Wallbrandt, P.; Carlstein, A.; Egelrud, T  
J. Biol. Chem. 269, 19420-19426, 1994  
A/Title: Cloning, expression, and characterization of stratum corneum chymotryptic enzyme  
A/Reference number: A53968; MUID:94308225; PMID:8034709  
A/Accession: A53968  
A/Status: preliminary  
A/Molecule type: mRNA  
A/Residues: 1-253 <HAN>  
A/Cross-references: GB:L33404; NID:G521214; PIDN:AAC37551.1; PID:G532504  
C/Genetics:  
A/Gene: GDS:PRSS6; SCCE  
A/Cross-references: GDB:1377730  
A/Map position: 7q35-q35  
C/Superfamily: trypsin; trypsin homology  
F:30-245/Domain: trypsin homology <TRI>

Query Match 100.0%; Score 47; DB 2; Length 253;  
Best Local Similarity 100.0%; Pred. No. 0.25;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 VLVNERWVL 9  
Db 58 VLVNERWVL 66

### RESULT 2

DBHU  
Complement factor D (EC 3.4.21.46) precursor [validated] - human (fragment)  
N/Alternate names: adipsin; C3 convertase activator  
C/Species: Homo sapiens (man)  
C/Date: 28-Aug-1985 #sequence\_revision 31-Dec-1992 #text\_change 08-Dec-2000  
C/Accession: A40197; A00936; A60571; S66645  
R/White, R.T.; Damm, D.; Hancock, N.; Rosen, B.S.; Lowell, B.B.; Usher, P.; Flier, J.S.  
J. Biol. Chem. 267, 9210-9213, 1992  
A/Title: Human adipsin is identical to complement factor D and is expressed at high leve  
A/Reference number: A40197; MUID:92250520; PMID:1374388  
A/Accession: A40197  
A/Molecule type: mRNA  
A/Residues: 1-246 <WHI>  
A/Cross-references: GB:M84526  
R/Niemann, M.A.; Brown, A.S.; Bennett, J.C.; Volanakis, J.E.  
Biochemistry 23, 2482-2486, 1984  
A/Title: Amino acid sequence of human D of the alternative complement pathway.  
A/Reference number: A00936; MUID:85000441; PMID:638466  
A/Accession: A00936

leucule type: protein  
 Residues: 19-44, 'G', 46-51, 'Q', 53-75, 'TH', 78, 'P', 80-93, 'XXXTIE', 90-172, 86-91, 185-235, 'A'; a few residues were assigned from the previously published sequence of Reid et al.  
 yta, T.; Oda, T.; Inagi, R.; Sugiyama, S.; Miyama, A.; Maeda, K.; Nakashima, I.; Ya  
 Immunol. 27, 637-644, 1990  
 title: Molecular and functional identification and purification of complement component  
 reference number: A60571; MUID:90370044; PMID:2395435  
 accession: A60571  
 clecule type: protein  
 Residues: 19-20, 'XX', 23-27, 'XX', 30-31, 'XX', 34, 'X', 36-40 <MIY>  
 alke, N.; Holtkamp, U.; Hoerl, W.H.; Teschesche, H.  
 S Lett. 371, 300-302, 1995  
 title: Inhibition of degranulation of human polymorphonuclear leukocytes by complement  
 reference number: S66645; MUID:96013156; PMID:755615  
 accession: S66645  
 title: preliminary  
 clecule type: protein  
 Residues: 19-44, 'C', 46-48 <BAL>  
 Comment: Factor D cleaves factor B when the latter is complexed with factor C3b, activ  
 Genetics:  
 Gene: GDB:DF  
 Cross-references: GDB:132645; OMIM:134350  
 Map position: Xpter-Xqter  
 Superfamily: trypsin; trypsin homology  
 Keywords: complement alternate pathway; hydrolase; plasma; serine proteinase  
 1-18/Domain: signal sequence #status predicted <SIG>  
 19-246/Product: complement factor D (fragment) #status experimental <MAT>  
 19-241/Domain: trypsin homology <TRY>  
 44-60, 141-207, 172-188, 197-222/Disulfide bonds: #status predicted  
 59, 105, 201/Active site: His, Asp, Ser #status predicted  
 Query Match 83.0%; Score 39; DB 1; Length 246;  
 Best Local Similarity 88.9%; Pred. No. 7.5;  
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 Y 1 VLVNERWVL 9  
 ||| |||||  
 b 47 VLVASRWVL 55  
 ||| |||||  
 RESULT 3  
 CPGC  
 issue kallikrein (EC 3.4.21.35), pancreatic - pig (tentative sequence)  
 V/Alternate names: glandular kallikrein; kininogenin  
 C/Species: Sus scrofa domestica (domestic pig)  
 C/Date: 24-Apr-1984 #sequence\_revision 31-Dec-1993 #text\_change 31-Mar-2000  
 C/Accession: A00938; A92895  
 R/Teschke, H.; Mair, G.; Godec, G.; Fiedler, F.; Ehrat, W.; Hirschauer, C.; Lemon, M.  
 Adv. Exp. Med. Biol. 120, 245-260, 1979  
 A/Title: The primary structure of porcine glandular kallikreins.  
 A/Reference number: A90015  
 A/Accession: A00938  
 A/Molecule type: protein  
 A/Residues: 1-49, 'GWL', 53-134, 'D', 136-156, 'H', 158, 'B', 160-224, 'B', 226-232 <TSC>  
 A/Note: the residue identified as 225-Asx is bound to carbohydrate; therefore, we have s  
 R/Bode, W.; Chen, Z.; Bartels, K.; Kutzbach, C.; Schmidt-Kastner, G.; Bartunik, H.  
 J. Mol. Biol. 164, 237-282, 1983  
 A/Title: Refined 2 angstrom X-ray crystal structure of porcine pancreatic kallikrein A.  
 C, structure and its comparison with bovine trypsin.  
 A/Reference number: A92895; MUID:8169107; PMID:6551452  
 A/Contents: X-ray crystallography, 2 angstroms  
 A/Accession: A92895  
 A/Molecule type: protein  
 A/Residues: 1-224, 'B', 226-232 <BOD>  
 C/Comment: The protein consists of two chains, A and B, held together by disulfide bonds  
 C/Comment: Tissue kallikreins cleave Met-Lys and Arg-Ser bonds in kininogen to release I  
 C/Superfamily: trypsin; trypsin homology  
 C/Keywords: glycoprotein; hydrolase; pancreas; serine proteinase  
 F/1-224/Domain: trypsin homology <TRY>  
 F/1-80/Product: tissue kallikrein chain A #status experimental <MPTA>  
 F/79-82/Region: autolysis loop  
 F/81-232/Product: tissue kallikrein chain B #status experimental <MPTB>  
 F/7-144, 26-42, 121-190, 155-169, 180-205/Disulfide bonds: #status experimental

F/41, 89, 184/Active site: His, Asp, Ser #status experimental  
 F/78/Binding site: carbohydrate (Asn) (covalent) #status experimental  
 F/225/Binding site: carbohydrate (Asn) (partial) #status experimental  
 Query Match 80.9%; Score 38; DB 1; Length 232;  
 Best Local Similarity 77.8%; Pred. No. 11;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 VLVNERWVL 9  
 ||| |||||  
 Db 29 VLVNPKWVL 37  
 ||| |||||  
 RESULT 4  
 I58559  
 neuropeptide - mouse  
 C/Species: Mus musculus (house mouse)  
 C/Date: 26-Jul-1996 #sequence\_revision 26-Jul-1996 #text\_change 21-Jul-2000  
 C/Accession: I58559  
 R/Chen, Z.L.; Yoshida, S.; Kato, K.; Momota, Y.; Suzuki, J.; Tanaka, T.; Ito, J.; Nishino  
 J. Neurosci. 15, 5088-5097, 1995  
 A/Title: Expression and activity-dependent changes of a novel limbic-serine protease gene  
 A/Reference number: I58559; MUID:95348817; PMID:7623137  
 A/Accession: I58559  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: mRNA  
 A/Residues: 1-260 <RES>  
 A/Cross-references: GB:D30785; NID:g1648847; PIDN:EAA06451.1; PID:g1020091  
 C/Superfamily: trypsin; trypsin homology  
 F/33-252/Domain: trypsin homology <TRY>  
 Query Match 80.9%; Score 38; DB 2; Length 260;  
 Best Local Similarity 77.8%; Pred. No. 12;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 VLVNERWVL 9  
 ||| |||||  
 Db 61 VLVGDRWVL 69  
 ||| |||||  
 RESULT 5  
 T25594  
 hypothetical protein C32E8.6 - Caenorhabditis elegans  
 C/Species: Caenorhabditis elegans  
 C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
 C/Accession: T25594  
 R/Gattung, S.  
 submitted to the EMBL Data Library, February 1997  
 A/Description: the sequence of C. elegans cosmid C32E8.  
 A/Reference number: Z20056  
 A/Accession: T25594  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: DNA  
 A/Residues: 1-417 <GAT>  
 A/Cross-references: EMBL:U88308; PIDN:AA842324.1; GSPDB:GN00019; CESP:C32E8.6  
 A/Experimental source: strain Bristol N2; clone C32E8  
 C/Genetics:  
 A/Gene: CESP:C32E8.6  
 A/Map position: 1  
 A/Introns: 22/3, 51/2, 92/1, 297/1, 354/1  
 Query Match 80.9%; Score 38; DB 2; Length 417;  
 Best Local Similarity 85.7%; Pred. No. 20;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VLVNERW 7  
 ||| |||||  
 Db 219 VLVNERW 225  
 ||| |||||  
 RESULT 6  
 T01779  
 trypsin (EC 3.4.21.4) - plaice

Species: Pleuronectes platessa (plaice)  
 Date: 19-Feb-1999 #sequence\_revision 19-Feb-1999 #text\_change 16-Jul-1999  
 Accession: T01779  
 Author: M.J.; George, S.G.  
 Submitted to the EMBL Data Library, March 1996  
 Reference number: Z14422  
 Accession: T01779  
 Status: preliminary; translated from GB/EMBL/DBJ  
 Molecule type: mRNA  
 Cross-references: EMBL:X56744; NID:gl213630; PID:g64240  
 Keywords: trypsin; trypsin homology  
 3-24/Domain: trypsin homology <TRY>  
 Query Match 78.7%; Score 37; DB 2; Length 250;  
 Best Local Similarity 66.7%; Pred. No. 18;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 1 VLVNERWVL 9  
 ||:||||  
 50 VLVNNQWVL 58  
 RESULT 7  
 S44184  
 chymotrypsin (EC 3.4.21.4) - plaice  
 N/Alternate names: Pleuronectes platessa (plaice)  
 C/Species: Pleuronectes platessa (plaice)  
 C/Date: 02-Dec-1993 #sequence\_revision 03-Aug-1995 #text\_change 18-Jul-1997  
 C/Accession: S31384  
 C/Author: M.J.; George, S.G.  
 Submitted to the EMBL Data Library, November 1990  
 Reference number: S30567  
 Accession: S31384  
 Status: preliminary  
 Molecule type: mRNA  
 Cross-references: EMBL:X56744  
 Residues: 1-250 <LEA>  
 Superfamily: trypsin; trypsin homology  
 Keywords: hydrolase; protein digestion; serine proteinase  
 F/23-24/Domain: trypsin homology <TRY>  
 Query Match 78.7%; Score 37; DB 2; Length 250;  
 Best Local Similarity 66.7%; Pred. No. 18;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 1 VLVNERWVL 9  
 ||:||||  
 50 VLVNNQWVL 58  
 RESULT 8  
 S44184  
 chymotrypsin (EC 3.4.21.1) 2 precursor - African malaria mosquito  
 N/Alternate names: chymotrypsin-like proteinase ANCHYM2  
 C/Species: Anopheles gambiae (African malaria mosquito)  
 C/Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 19-May-2000  
 C/Accession: S44184; S45130  
 C/Author: H.M.; Catteruccia, F.; Crisanti, A.  
 Submitted to the EMBL Data Library, April 1994  
 Reference number: S44184  
 Accession: S44184  
 Status: preliminary  
 Molecule type: DNA  
 Cross-references: EMBL:Z32645; NID:g474026; PIDN:CAA83567.1; PID:g474027  
 Keywords: hydrolase; protein digestion; serine proteinase  
 F/23-24/Domain: trypsin homology <TRY>  
 Query Match 78.7%; Score 37; DB 2; Length 250;  
 Best Local Similarity 66.7%; Pred. No. 18;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 1 VLVNERWVL 9  
 ||:||||  
 50 VLVNNQWVL 58  
 RESULT 9  
 S44184  
 chymotrypsin (EC 3.4.21.1) 1 precursor - African malaria mosquito  
 N/Alternate names: chymotrypsin-like proteinase ANCHYM1  
 C/Species: Anopheles gambiae (African malaria mosquito)  
 C/Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 19-May-2000  
 C/Accession: S44184; S44185  
 C/Author: H.M.; Catteruccia, F.; Crisanti, A.  
 Submitted to the EMBL Data Library, April 1994  
 Reference number: S44184  
 Accession: S44185  
 Status: preliminary  
 Molecule type: DNA  
 Cross-references: EMBL:Z32645; NID:g474026; PIDN:CAA83568.1; PID:g474028  
 Keywords: hydrolase; protein digestion; serine proteinase  
 F/23-24/Domain: trypsin homology <TRY>  
 Query Match 78.7%; Score 37; DB 1; Length 259;  
 Best Local Similarity 75.0%; Pred. No. 19;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 2 VLVNERWVL 9  
 ||:||||  
 63 LLNDRWVL 70  
 RESULT 10  
 T05892  
 hypothetical protein FEH1.110 - Arabidopsis thaliana  
 C/Species: Arabidopsis thaliana (mouse-ear cress)  
 C/Date: 09-Apr-1999 #sequence\_revision 09-Apr-1999 #text\_change 13-Aug-1999  
 C/Accession: T05892  
 R/Bevan, M.; Brandt, P.; Dose, S.; Jarke, D.; Scharfe, M.; Schon, O.; Bancroft, I.; Mewe,

A/Cross-references: EMBL:Z18888; NID:g509416; PIDN:CAA79325.1; PID:g509417  
 C/Genetics:  
 A/Introns: 78/2; 209/3  
 C/Superfamily: trypsin; trypsin homology  
 C/Keywords: hydrolase; protein digestion; serine proteinase  
 F/1-17/Domain: signal sequence #status predicted <SIG>  
 F/18-32/Domain: activation peptide #status predicted <PRO>  
 F/33-259/Product: chymotrypsin 2 #status predicted <MAT>  
 F/33-259/Product: chymotrypsin 1 #status predicted <MAT>  
 F/59-75/Domain: trypsin homology <TRY>  
 F/59-75/182-198/208-232/disulfide bonds: #status experimental  
 F/74,119,212/Active site: His, Asp, Ser #status predicted  
 Query Match 78.7%; Score 37; DB 1; Length 259;  
 Best Local Similarity 75.0%; Pred. No. 19;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 2 VLVNERWVL 9  
 ||:||||  
 63 LLNDRWVL 70  
 RESULT 9  
 S44129  
 chymotrypsin (EC 3.4.21.1) 1 precursor - African malaria mosquito  
 N/Alternate names: chymotrypsin-like proteinase ANCHYM1  
 C/Species: Anopheles gambiae (African malaria mosquito)  
 C/Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 19-May-2000  
 C/Accession: S49129; S44185  
 C/Author: H.M.; Catteruccia, F.; Crisanti, A.  
 Submitted to the EMBL Data Library, November 1992  
 Reference number: S35412  
 Accession: S49129  
 Status: preliminary  
 Molecule type: mRNA  
 Residues: 1-259 <MUE>  
 Cross-references: EMBL:Z18887; NID:g509414; PIDN:CAA79325.1; PID:g509415  
 Experimental source: strain Suakoko  
 R/Mueller, H.M.; Catteruccia, F.; Crisanti, A.  
 Submitted to the EMBL Data Library, April 1994  
 Description: An Anopheles gambiae locus containing the sequences of two closely related  
 A/Reference number: S44184  
 Accession: S44185  
 Molecule type: DNA  
 Residues: 1-259 <MUE>  
 Cross-references: EMBL:Z32645; NID:g474026; PIDN:CAA83568.1; PID:g474028  
 Experimental source: strain Suakoko  
 C/Genetics:  
 A/Introns: 78/2; 209/3  
 C/Superfamily: trypsin; trypsin homology  
 C/Keywords: hydrolase; protein digestion; serine proteinase  
 F/1-17/Domain: signal sequence #status predicted <SIG>  
 F/18-32/Domain: activation peptide #status predicted <PRO>  
 F/33-259/Product: chymotrypsin 1 #status predicted <MAT>  
 F/33-259/Product: chymotrypsin 2 #status predicted <MAT>  
 F/59-75/Domain: trypsin homology <TRY>  
 F/59-75/182-198/208-232/disulfide bonds: #status experimental  
 F/74,119,212/Active site: His, Asp, Ser #status predicted  
 Query Match 78.7%; Score 37; DB 1; Length 259;  
 Best Local Similarity 75.0%; Pred. No. 19;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 2 VLVNERWVL 9  
 ||:||||  
 63 LLNDRWVL 70  
 RESULT 10  
 T05892  
 hypothetical protein FEH1.110 - Arabidopsis thaliana  
 C/Species: Arabidopsis thaliana (mouse-ear cress)  
 C/Date: 09-Apr-1999 #sequence\_revision 09-Apr-1999 #text\_change 13-Aug-1999  
 C/Accession: T05892  
 R/Bevan, M.; Brandt, P.; Dose, S.; Jarke, D.; Scharfe, M.; Schon, O.; Bancroft, I.; Mewe,

itted to the Protein Sequence Database, April 1998

Accession: Z15456  
 Molecule type: DNA  
 Residues: 1-1421 <BEV>  
 Cross-references: EMBL:AL021684  
 Experimental source: cultured Columbia, BAC clone F6H11  
 Annotations:  
 ap position: 5  
 ntions: 72/3; 142/3; 227/2; 344/3; 372/3; 941/2; 1016/3; 1038/3; 1055/1; 1095/2; 1159  
 Note: F6H11.110

Query Match 76.6%; Score 37; DB 2; Length 1421;  
 Best Local Similarity 66.7%; Pred. No. 1.1e+02;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

1 VLVNERWVL 9  
 ||| |||||  
 1208 VLVNERWVL 1216

# RESULT 11

hypothetical protein XF0279 [imported] - Xylella fastidiosa (strain 945c)  
 Species: Xylella fastidiosa  
 Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000  
 Accession: F82825  
 Note: anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequencing  
 Title: The genome sequence of the plant pathogen Xylella fastidiosa.  
 Reference number: A82515; MUID:20365717; PMID:10910347  
 Note: for a complete list of authors see reference number A59328 below  
 Accession: F82825  
 Status: preliminary  
 Molecule type: DNA  
 Residues: 1-72 <Sim>  
 Cross-references: GB:A8003981; GB:A8003849; NID:99105093; PIDN:AF83092.1; GSPDB:GN001  
 Experimental source: strain 945c  
 Authors: Simpson, A.J.G.; Reinach, P.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A  
 rriores, M.R.S.; Bueno, M.R.P.; Camargo, L.B.A.; Carraro, D.M.; Carret, H  
 as Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.  
 Submitted to GenBank, June 2000  
 Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm  
 J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Klieger, J.E.; Kuramae, E.E.; Laig  
 chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E  
 A.; Authors: Martins, E.M.F.; Matekuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;  
 P.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.  
 Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak  
 A.; Authors: da Silva, A.C.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir  
 M.; Tsunako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z

Query Match 76.6%; Score 36; DB 2; Length 72;  
 Best Local Similarity 62.5%; Pred. No. 7.4;  
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

1 VLVNERWV 8  
 ||| |||||  
 50 ILVNERWV 57

# RESULT 12

A45161  
 serine proteinase (EC 3.4.21.-) RNK-Met-1 - rat (fragment)  
 Species: Rattus norvegicus (Norway rat)  
 Date: 26-May-1994 #sequence\_revision 26-May-1994 #text\_change 22-Jun-1999  
 Accession: A45161  
 Note: R:Smeth, M.J.; Wilttrout, T.; Trapani, J.A.; Ottaway, K.S.; Sowder, R.; Henderson, L.E.;  
 J. Biol. Chem. 267, 24418-24425, 1992  
 A>Title Purification and cloning of a novel serine protease, RNK-Met-1, from the granu

A:Reference number: A45161; MUID:93077530; PMID:1447189  
 A:Accession: A45161  
 A>Status: Preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-258 <SMV>  
 A:Cross-references: GB:L05175; NID:9206691; PIDN:AAA42056.1; PID:9206692  
 C:Superfamily: trypsin; trypsin homology  
 C:Keywords: hydrolase; serine proteinase  
 F:21-245/Domain: trypsin homology <TRI>

Query Match 76.6%; Score 36; DB 2; Length 258;  
 Best Local Similarity 66.7%; Pred. No. 29;  
 Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

1 VLVNERWVL 9  
 ||| |||||  
 49 VLVHCKWVL 57

# RESULT 13

hypothetical protein A21R - variola major virus  
 Species: variola major virus  
 Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 21-Jul-2000  
 Accession: T28563  
 Note: R:Massung, R.F.; Esposito, J.J.; Liu, L.I.; Qi, J.; Utterback, T.R.; Knight, J.C.; Aubin,  
 Nature 366, 748-751, 1993  
 A>Title: Potential virulence determinants in terminal regions of variola smallpox virus  
 A:Reference number: Z20488; MUID:94088747; PMID:8264798  
 Accession: T28563  
 Status: preliminary; translated from GB/EMBL/DBJ  
 Molecule type: DNA  
 Residues: 1-426 <MAS>  
 Cross-references: EMBL:L22579; NID:9623595; PIDN:AAA60873.1; PID:g439043  
 Experimental source: strain Bangladesh-1975  
 Superfamily: vaccinia virus probable 49.1K protein

Query Match 76.6%; Score 36; DB 2; Length 426;  
 Best Local Similarity 55.6%; Pred. No. 45;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

1 VLVNERWVL 9  
 ||| |||||  
 130 VLVNRRWIM 138

# RESULT 14

A21R protein - variola virus (strain India-1967)  
 Species: variola virus  
 Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 23-Mar-2001  
 Accession: C36850  
 Note: submitted to GenBank, November 1992  
 Reference number: A36859  
 Accession: C36850  
 Status: preliminary  
 Molecule type: DNA  
 Residues: 1-426 <BLI>  
 Cross-references: GB:X69198; NID:9456758; PIDN:CAA49065.1; PID:g297303  
 Superfamily: vaccinia virus probable 49.1K protein

Query Match 76.6%; Score 36; DB 2; Length 426;  
 Best Local Similarity 55.6%; Pred. No. 49;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

1 VLVNERWVL 9  
 ||| |||||  
 130 VLVNRRWIM 138

# RESULT 15

C72166

id Mar 3 07:49:12 2004

us-09-905-083 3412

protein - variola minor virus (strain Garcia-1966)  
species: variola minor virus  
date: 24-Nov-1999 #sequence\_revision 24-Nov-1999 #text\_change 20-Jun-2000  
accession: C72166  
submitter: S.N.; Totmenin, A.V.; Gutorov, V.V.; Saifonov, P.F.; Massung, R.F.; Lopar  
submitted to GenBank, March 1998  
description: Analysis of the complete coding sequence of DNA of alastrim variola minor  
reference number: A72150  
accession: C72166  
status: preliminary  
molecule type: DNA  
residues: 1-426 <SHC>  
cross-references: GB:Y16780; NID:G5830555; PIDN:CAB54724.1; PID:G5830685  
experimental source: strain Garcia-1966  
genetics:  
gene: A22R  
superfamily: vaccinia virus probable 49.1K protein  
Query Match 76.6%; Score 36; DB 2; Length 426;  
Best Local Similarity 55.6%; Pred. No. 49;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
1 VLVNERWVL 9  
130 VLLNRRWIM 138

Search completed: March 1, 2004, 17:36:28  
Job time : 10.3333 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

protein - protein search, using sw model

on: March 1, 2004, 17:20:41 ; Search time 30.3333 seconds  
(without alignments)  
93.615 Million cell updates/sec

US-09-905-083-34

Effect score: 47

quence: 1 VLVNERWYL 9

oring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

arched: 1017041 seqs, 315518202 residues

al number of hits satisfying chosen parameters: 1017041

imum DB seq length: 0

iximum DB seq length: 2000000000

st-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

atabase : SPTREMBL\_25.\*

- 1: sp\_archae.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phase.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_rvirus.\*
- 16: sp\_bacteriap.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	47	100.0	253	4 Q8NSN9	Q8NSN9 homo sapien
2	40	85.1	455	10 Q9FPR7	Q9FPR7 arabidopsis
3	39	83.0	100	6 Q8MYV8	Q8MYV8 bos taurus
4	39	83.0	249	13 Q8JFQ7	Q8JFQ7 gadus morhu
5	38	80.9	223	5 Q9X158	Q9X158 ctenocephal
6	38	80.9	234	13 Q7S2E2	Q7S2E2 agkistrodon
7	38	80.9	262	5 Q7Z0G5	Q7Z0G5 phlebotomus
8	38	80.9	313	5 Q8T3B8	Q8T3B8 caenorhabdi
9	38	80.9	377	10 Q9SNY8	Q9SNY8 solanum tub
10	38	80.9	418	10 Q9SNY9	Q9SNY9 solanum tub
11	38	80.9	461	5 P81123	P81123 caenorhabdi
12	37	78.7	233	13 Q9P5L1	Q9P5L1 agkistrodon
13	37	78.7	234	13 Q8UUL1	Q8UUL1 agkistrodon
14	37	78.7	243	5 Q9VEM7	Q9VEM7 drosophila
15	37	78.7	249	13 Q9W6K0	Q9W6K0 notonthenia
16	37	78.7	249	13 Q92046	Q92046 dissoastichu

17 37 78.7 250 13 Q93265  
18 37 78.7 254 5 Q97098  
19 37 78.7 255 5 Q97100  
20 37 78.7 257 13 Q9PTL3  
21 37 78.7 257 13 Q9YJ39  
22 37 78.7 258 5 Q9XYS3  
23 37 78.7 258 13 Q7S2E1  
24 37 78.7 259 13 Q8UUK2  
25 37 78.7 262 13 Q8QHK3  
26 37 78.7 344 13 Q9M6J9  
27 37 78.7 348 4 Q8WNS5  
28 37 78.7 675 13 Q9W6J8  
29 37 78.7 1176 16 Q981K4  
30 36 76.6 72 16 Q9PGL9  
31 36 76.6 111 4 Q9UKR2  
32 36 76.6 259 5 Q45048  
33 36 76.6 295 16 Q98N76  
34 36 76.6 299 5 Q9VS87  
35 36 76.6 345 11 Q923Y5  
36 36 76.6 350 16 Q98JL3  
37 36 76.6 426 12 Q97288  
38 36 76.6 426 12 Q9JF80  
39 36 76.6 426 12 Q8V2N4  
40 36 76.6 426 12 Q8VAV6  
41 36 76.6 426 12 Q8QMT2  
42 36 76.6 426 12 Q8JL93  
43 36 76.6 426 12 Q89166  
44 36 76.6 426 12 Q80HV4  
45 36 76.6 426 12 Q80DV4

## ALIGNMENTS

## RESULT 1

Q8NSN9 PRELIMINARY; PRT; 253 AA.  
AC Q8NSN9;  
DT 01-OCT-2002 (trEMBLrel. 22, Created)  
DT 01-OCT-2002 (trEMBLrel. 22, Last sequence update)  
DT 01-OCT-2003 (trEMBLrel. 25, Last annotation update)  
DE Kallikrein 7 (chymotryptic, stratum corneum).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Skin;  
RA Strausberg R;  
RL Submitted (JUN-2002) to the EMBL/GenBank/DDBJ databases.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.  
DR EMBL; BC032005; AAH32005.1; -.  
DR GO; GO:0004263; F:chymotrypsin activity; IEA.  
DR GO; GO:0008233; F:peptidase activity; IEA.  
DR GO; GO:0004295; F:trypsin activity; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR InterPro; IPR009003; Cys Ser trypsin.  
DR InterPro; IPR001254; Peptidase\_S1.  
DR InterPro; IPR001314; Peptidase\_S1.  
DR Pfam; PF00089; trypsin; 1.  
DR PRINTS; PR00722; CHYMOTRYPSIN.  
DR SMART; SM00020; TRY SP; 1.  
DR PROSITE; PSS0240; TRYPSIN\_DOM; 1.  
DR PROSITE; PS0134; TRYPSIN\_HIS; 1.  
DR PROSITE; PS0135; TRYPSIN\_SER; 1.  
KW Hydrolase; Protease; Serine protease.  
SQ SEQUENCE 253 AA; 27608 MW; 2D68B6A41B22A668 CRC64;

Query Match 100.0%; Score 47; DB 4; Length 253;  
Best Local Similarity 100.0%; Pred. No. 0.64; 0; Indels 0; Caps 0;  
Matches 9; Conservative 0; Mismatches 0;

1 VLVNERWVL 9  
|||||  
58 VLVNERWVL 66

## RESULT 2

Q9FFR7 PRELIMINARY; PRT; 455 AA.

01-MAR-2001 (TRENBLrel. 16, Created)  
01-MAR-2001 (TRENBLrel. 16, Last sequence update)  
01-OCT-2003 (TRENBLrel. 25, Last annotation update)  
Genomic DNA, chromosome 5, pl clone:MBX23.  
Arabidopsis thaliana (Mouse-ear cross).  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
eurosids II; Brassicales; Brassicaceae; Arabidopsi.  
NCBI\_TaxID=3702;  
[1]  
SEQUENCE FROM N.A.  
STRAINS=Columbia;  
MEDLINE=97471969; PubMed=9330910;  
Sato S., Kotani H., Nakamura Y., Kaneko T., Asamizu E., Fukami M.,  
Miyaajima N., Tabata S.;  
"Structural analysis of Arabidopsis thaliana chromosome 5. I. Sequence  
features of the 1.6 Mb regions covered by twenty physically assigned  
pl clones";  
DNA Res. 4:215-230(1997).  
EMBL; AB005233; BAB11469.1; -.  
InterPro; IPR001810; F-box.  
Fam; PF00646; F-box; 1.  
SEQUENCE 455 AA; 52305 MW; 763F5D5278D22F13 CRC64;

Query Match 85.1%; Score 40; DB 10; Length 455;  
Best Local Similarity 75.0%; Pred. No. 26;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

2 LVNERWVL 9  
|||||  
104 LVNERWVL 111

## RESULT 3

Q9MYV8 PRELIMINARY; PRT; 100 AA.

01-OCT-2000 (TRENBLrel. 15, Created)  
01-OCT-2000 (TRENBLrel. 15, Last sequence update)  
01-OCT-2003 (TRENBLrel. 25, Last annotation update)  
Haptoglobin (Fragment).  
HP.  
Bos taurus (Bovine).  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
Bovidae; Bovinae; Bos.  
NCBI\_TaxID=9913;  
[1]  
SEQUENCE FROM N.A.  
TISSUE=Oviduct;  
Lavery K.S., Gabler C., Killian G.J.;  
"Expression and localization of haptoglobin in the bovine female  
reproductive tract";  
Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
EMBL; AJ271156; CAC00531.1; -.  
HSP; P00734; IBAB.  
MEROPS; S01.972; -.  
GO; GO:0008233; P:peptidase activity; IEA.  
GO; GO:0004295; F:trypsin activity; IEA.  
GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
InterPro; IPR009003; Cys Ser trypsin.  
InterPro; IPR001254; Peptidase\_S1.  
Fam; PF00089; trypsin; 1.

DR PROSITE; PS0240; TRYPSIN\_DOM; 1.  
KW Hydrolase; Protease; Serine protease.  
FT NON\_TER 1  
FT NON\_TER 100  
SQ SEQUENCE 100 AA; 11239 MW; FFF6A5AB43CED9 CRC64;

Query Match 83.0%; Score 39; DB 6; Length 100;  
Best Local Similarity 75.0%; Pred. No. 8.2;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LVNERWVL 9  
|||||  
22 LVNERWVL 29

## RESULT 4

Q8JFQ7 PRELIMINARY; PRT; 249 AA.  
AC Q8JFQ7;  
DT 01-OCT-2002 (TRENBLrel. 22, Created)  
DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)  
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)  
DE Trypsinogen Y precursor (Fragment).  
GN TRPY GENE.  
OS Gadus morhua (Atlantic cod).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
OC Acanthomorpha; Paracanthopterygii; Gadiformes; Gadidae; Gadus.  
OX NCBI\_TaxID=8049;  
[1]  
SEQUENCE FROM N.A.  
TISSUE=Pyloric caecum;  
RC Gudmundsdottir A.;  
RA Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.  
[2]  
SEQUENCE FROM N.A.  
RC TISSUE=Pyloric caecum;  
RX PubMed=10612685;  
RA Spilliaert R.;  
RT "Atlantic Cod Trypsin Y - Member of a Novel Trypsin Group.";  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.  
DR EMBL; AJ459311; CAD30563.1; -.  
DR GO; GO:0004263; F:chymotrypsin activity; IEA.  
DR GO; GO:0008233; F:peptidase activity; IEA.  
DR GO; GO:0004295; F:trypsin activity; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR InterPro; IPR009003; Cys Ser trypsin.  
DR InterPro; IPR001254; Peptidase\_S1.  
DR InterPro; IPR001314; Peptidase\_S1A.  
DR Pfam; PF00089; trypsin; 1.  
DR PRINTS; PR00722; CHYMOTRYPSIN.  
DR SMART; SM00020; TRYP\_SPC; 1.  
DR PROSITE; PS0240; TRYPSIN\_DOM; 1.  
DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
KW Hydrolase; Protease; Serine protease; Signal.  
FT SIGNAL 1 15 POTENTIAL.  
FT CHAIN 23 >249 TRYPSINOGEN Y.  
FT NON\_TER 249 249  
SQ SEQUENCE 249 AA; 27400 MW; 8FB98462CEDBEFC9 CRC64;

Query Match 83.0%; Score 39; DB 13; Length 249;  
Best Local Similarity 66.7%; Pred. No. 21;  
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLVNERWVL 9  
|||||  
50 VLVNQWVL 58

## RESULT 5

Q9XY58 PRELIMINARY; PRT; 223 AA.



Q9XY58;  
 01-NOV-1999 (TrEMBLrel. 12, Created)  
 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 Chymotrypsin-like serine protease (Fragment).  
 SP-4  
 Ctenocephalides felis (Cat flea).  
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 Neoptera; Endopterygota; Siphonaptera; Pulicidae; Pulicinae;  
 Ctenocephalides.  
 NCBI\_TaxID=7515;  
 [1]  
 SEQUENCE FROM N.A.  
 A Gaines P.J., Samson C.M., Ruehlow K.E., Stiegler G.L.;  
 "Cloning of a family of serine protease genes from the cat flea  
 Ctenocephalides felis";  
 Insect Mol. Biol. 81:11-22(1999).  
 -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.  
 EMBL; AF053916; AAD21836.1; -.  
 R HSP; P00766; 1CHG.  
 R GO; GO:0004263; F:chymotrypsin activity; IEA.  
 R GO; GO:0008233; F:peptidase activity; IEA.  
 R GO; GO:0004295; F:trypsin activity; IEA.  
 R GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
 R InterPro; IPR009003; Cys Ser trypsin.  
 R InterPro; IPR001254; Peptidase S1.  
 R InterPro; IPR001314; Peptidase\_S1A.  
 R Pfam; PF00089; trypsin; 1.  
 R PRINTS; PR00722; CHYMOTRYPSIN.  
 R SMART; SM00020; ITP\_SPC; 1.  
 R PROSITE; PS00240; TRYPSIN\_DOM; 1.  
 R PROSITE; PS00134; TRYPSIN\_HIS; 1.  
 R PROSITE; PS00135; TRYPSIN\_SER; 1.  
 W Hydrolase; Protease; Serine protease.  
 T NON\_TER 1  
 Q SEQUENCE 223 AA; 24844 MW; C34EBE9455DD7DE9 CRC64;

Query Match 80.9%; Score 38; DB 5; Length 223;  
 Best Local Similarity 75.0%; Pred. No. 29;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Y 2 LVNERWVL 9  
 : : : : :  
 b 35 ILNERWVL 42  
 : : : : :  
 RESULT 6  
 7SZE2 PRELIMINARY; PRT; 234 AA.  
 D Q7SZE2 (TrEMBLrel. 25, Created)  
 C Q7SZE2 (TrEMBLrel. 25, Last sequence update)  
 T 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 T 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 E Thrombin-like enzyme defibrase (Fragment).  
 S Agkistrodon caliginosus (Korean viper) (Gloydius ussuriensis).  
 C Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 C Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;  
 C Viperidae; Crotalinae; Gloydius.  
 X NCBI\_TaxID=35671;  
 [1]  
 N SEQUENCE FROM N.A.  
 P TISSUE=Venom gland;  
 A Sun D.-J., Yang T.-S.;  
 J Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.  
 L EMBL; AY204242; AAP20637.1; -.  
 T NON\_TER 1  
 Q SEQUENCE 234 AA; 25351 MW; 0124C1485525F560 CRC64;  
 Query Match 80.9%; Score 38; DB 13; Length 234;  
 Best Local Similarity 87.5%; Pred. No. 31;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Y 2 LVNERWVL 9

Db 30 LVNEEWVL 37  
 : : : : :  
 RESULT 7  
 Q7ZOG5 PRELIMINARY; PRT; 262 AA.  
 AC Q7ZOG5 (TrEMBLrel. 25, Created)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)  
 DE Chymotrypsin.  
 GN CHYM1.  
 OS Phlebotomus papatasi.  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Nematocera; Psychodoidea;  
 OC Psychodidae; Phlebotomus.  
 OK NCBI\_TaxID=29031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Midgut;  
 RA Ramalho-Ortigao M., Kamhawi S., Rowton E., Sacks D.L., Ribeiro J.M.,  
 RA Valenzuela J.G.;  
 RT "Cloning and characterization of trypsins and chymotrypsins-like  
 RT proteases from the midgut of the sand fly vector Phlebotomus  
 RT papatasi";  
 RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AY128106; AAM96938.1; -.  
 SQ SEQUENCE 262 AA; 28863 MW; 31F11012206A65D4 CRC64;  
 Query Match 80.9%; Score 38; DB 5; Length 262;  
 Best Local Similarity 75.0%; Pred. No. 35;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LVNERWVL 9  
 : : : : :  
 Db 57 ILNERWVL 64  
 : : : : :

RESULT 8  
 Q8T3B8 PRELIMINARY; PRT; 313 AA.  
 AC Q8T3B8 (TrEMBLrel. 21, Created)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
 DE Hypothetical protein.  
 DE C32E8.6.  
 GN Caenorhabditis elegans.  
 OS Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RX MEDLINE=99069613; PubMed=9851916;  
 RA Waterston R.;  
 RT "Genome sequence of the nematode C. elegans: a platform for  
 RT investigating biology. The C. elegans Sequencing Consortium.";  
 RL Science 282:2012-2018(1998).  
 [2]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=Bristol N2;  
 RC Gattung S.;  
 RA "The sequence of C. elegans cosmid C32E8.";  
 RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.  
 [3]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=Bristol N2;  
 RC Waterston R.;  
 RA Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; U88308; AAM15557.1; -.  
 DR WormPep; C32E8.6b; CE30371.

```

GO; GO:0003824; F:catalytic activity; IEA.
GO; GO:0008152; P:metabolism; IEA.
InterPro; IPR000873; AMP-binding.
Pfam; PF00501; AMP-binding; 1.
PROSITE; PS00455; AMP_BINDING; 1.
SEQUENCE 313 AA; 35608 MW; 4372ED6478E434A CRC64;
Query Match 80.9%; Score 38; DB 5; Length 313;
Best Local Similarity 85.7%; Pred.No. 42;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

1 VLVNERW 7
|||:|||||
219 VLVNERW 225

RESULT 9
9SNY8 PRELIMINARY; PRT; 377 AA.
9SNY8;
01-MAY-2000 (TREMBLrel. 13, Created)
01-MAY-2000 (TREMBLrel. 13, Last sequence update)
01-OCT-2003 (TREMBLrel. 25, Last annotation update)
Branched-chain amino acid aminotransferase (EC 2.6.1.42).
BCAT2.
Solanum tuberosum (Potato).
Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; asterids;
lamiales; Solanales; Solanaceae; Solanum.
NCBI_TaxID=4113;
[1]
SEQUENCE FROM N.A.
STRAIN=cv. Russet Burbank;
Campbell M.A., Patel J.K., Meyers J.L., Myrick L.C.;
"Isolation of a functional gene encoding for branched-chain amino acid
aminotransferase from potato.";
Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
EMBL; AF193845; AAF07192.1; -.
HSSP; P00510; 1A3G.
GO; GO:0004084; F:branched-chain-amino-acid transaminase acti. .; IEA.
GO; GO:0016740; F:transferase activity; IEA.
GO; GO:0009081; P:branched chain family amino acid metabolism; IEA.
GO; GO:0008152; P:metabolism; IEA.
InterPro; IPR001544; Aminotrans_IV.
InterPro; IPR005786; B_amino_transII.
Pfam; PF01063; aminotran_4; 1.
ProbDom; PD001961; Aminotran_4; 1.
TIGRFAMs; TIGR01123; ilve_II; 1.
Aminotransferase; Transferase.
KW SEQUENCE 377 AA; 40924 MW; 8CC8460A5F15B646 CRC64;
SEQUENCE 377 AA; 40924 MW; 8CC8460A5F15B646 CRC64;
Query Match 80.9%; Score 38; DB 10; Length 377;
Best Local Similarity 75.0%; Pred.No. 51;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

1 VLVNERW 8
|||:|||||
147 VLVNERW 154

RESULT 10
9SNY9 PRELIMINARY; PRT; 418 AA.
9SNY9;
01-MAY-2000 (TREMBLrel. 13, Created)
01-MAY-2000 (TREMBLrel. 13, Last sequence update)
01-OCT-2003 (TREMBLrel. 25, Last annotation update)
Branched-chain amino acid aminotransferase (EC 2.6.1.42).
BCAT1.
Solanum tuberosum (Potato).
Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;

```

```

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC lamiales; Solanales; Solanaceae; Solanum.
NCBI_TaxID=4113;
[1]
SEQUENCE FROM N.A.
STRAIN=cv. Russet Burbank;
Campbell M.A., Patel J.K., Meyers J.L., Myrick L.C.;
"Isolation of a functional gene encoding for branched-chain amino acid
aminotransferase from potato.";
Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
EMBL; AF193845; AAF07191.1; -.
HSSP; P00510; 1A3G.
DR DR GO; GO:0004084; F:branched-chain-amino-acid transaminase acti. .; IEA.
DR DR GO; GO:0016740; F:transferase activity; IEA.
DR DR GO; GO:0009081; P:branched chain family amino acid metabolism; IEA.
DR DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR001544; Aminotrans_IV.
DR InterPro; IPR005786; B_amino_transII.
DR Pfam; PF01063; aminotran_4; 1.
DR ProbDom; PD001961; Aminotran_4; 1.
DR TIGRFAMs; TIGR01123; ilve_II; 1.
KW Aminotransferase; Transferase.
SQ SEQUENCE 418 AA; 45296 MW; E6EA6C2B3A58C814 CRC64;
Query Match 80.9%; Score 38; DB 10; Length 418;
Best Local Similarity 75.0%; Pred.No. 57;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

1 VLVNERW 8
|||:|||||
188 VLVNERW 195

RESULT 11
91123 PRELIMINARY; PRT; 461 AA.
91123;
AC P91123;
DT 01-MAY-1997 (TREMBLrel. 03, Created)
DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN C32E8.6.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
NCBI_TaxID=6239;
[1]
SEQUENCE FROM N.A.
STRAIN=Bristol N2;
RC STRAIN=Bristol N2;
RX MEDLINE=90069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
[2]
SEQUENCE FROM N.A.
STRAIN=Bristol N2;
RC STRAIN=Bristol N2;
RA Gattung S.;
RT "The sequence of C. elegans cosmid C32E8.";
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
[3]
SEQUENCE FROM N.A.
STRAIN=Bristol N2;
RC STRAIN=Bristol N2;
RA Waterston R.;
RT "Direct Submission.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; U88308; AAB42324.2; -.
DR PIR; T25594; T25594.
DR HSP; P08659; ILCI.
DR WormRep; C32E8.6a; CE27814.
DR DR GO; GO:0003824; F:catalytic activity; IEA.
DR DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR000873; AMP-bind.

```

```

3 Pfam: PF00501; AMP-binding; 2.
3 PROSITE; PS00455; AMP_BINDING; 1.
4 Hypothetical protein_519202; 8662F7486A0B57A1 CRC64;
2 SEQUENCE 461 AA; 51902 MW; 8662F7486A0B57A1 CRC64;

Query Match      80.9%; Score 38; DB 5; Length 461;
Best Local Similarity 85.7%; Pred. No. 63;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Y 1 VLVNERN 7
  ||:|||||
b 219 VLVNERN 225

RESULT 12
9PT51 PRELIMINARY; PRT; 233 AA.
D Q9PT51
T 01-MAY-2000 (TRENBLrel. 13, Created)
T 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
T 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
E Beta-fibrinogenase (Fragment).
N BREVINASE.
S Agkistrodon halys blomhoffii (Mamushi) (Gloydus blomhoffii).
C Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
X Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
X Viperidae; Crotalinae; Agkistrodon.
X NCBI_TaxID=61300;
[1]
N SEQUENCE FROM N.A.
P MEDLINE=20302322; PubMed=10845699;
A Lee J.W., Park W.;
T "cDNA cloning of brevinase, a heterogeneous two-chain fibrinolytic
T enzyme from Agkistrodon blomhoffii reaction."
T Hybridization-polymerase chain reaction."
T Arch. Biochem. Biophys. 377:234-240(2000).
L C -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
D EMBL; AJ243757; CAB65936.1; -.
D HSSP; P00763; IDPO.
D MEROPS; S01.185; -.
D GO; GO:0004263; F:chymotrypsin activity; IEA.
D GO; GO:0008233; F:peptidase activity; IEA.
D GO; GO:0004295; F:trypsin activity; IEA.
D GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
D InterPro; IPR009003; Cys Ser trypsin.
D InterPro; IPR001254; Peptidase_S1.
D Pfam; PF00089; trypsin; 1.
D PRINTS; PR00722; CHYMOTRYPSIN.
D SMART; SM00020; TRYPSIN_DOM; 1.
D PROSITE; PS00240; TRYPSIN_HIS; 1.
D PROSITE; PS00134; TRYPSIN_HIS; 1.
D PROSITE; PS00135; TRYPSIN_SER; 1.
D Hydrolyase; Protease; Serine protease.
D NON TER 1
D SEQUENCE 233 AA; 25725 MW; 1676DC5AFA0B5A42 CRC64;

Query Match      78.7%; Score 37; DB 13; Length 233;
Best Local Similarity 75.0%; Pred. No. 48;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

2y 2 LVNERNWL 9
  ||:|||||
Db 29 LVNERNWL 36

RESULT 13
28UJ1 PRELIMINARY; PRT; 234 AA.
AC Q8UJ1
D Q8UJ1
T 01-MAR-2002 (TRENBLrel. 20, Created)
T 01-MAR-2002 (TRENBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)

```

```

DE Thrombin-like enzyme ussurase (Fragment).
OS Agkistrodon caliginosus (Korean viper) (Gloydus ussuriensis).
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
CC Viperidae; Crotalinae; Gloydus.
OX NCBI_TaxID=35671;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Poison gland;
RA Daolin D., Xiaodong D., Wenfang W., Anguo L., Mei X.;
RT "Cloning and Sequence Analysis of cDNAs for Thrombin-like Enzymes,
RT Ussurin and Ussurase, From Agkistrodon halys ussuriensis Snake
RT Venom."
RL She Zhi 13:1-9(2001).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
DR EMBL; AF44251; AAL48222.1; -.
DR HSSP; P00761; IANI.
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys Ser trypsin.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; TRYPSIN_DOM; 1.
DR PROSITE; PS00240; TRYPSIN_HIS; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
D Hydrolyase; Protease; Serine protease.
D NON TER 1
D SEQUENCE 234 AA; 26147 MW; A1EBE6E2FA29920 CRC64;

Query Match      78.7%; Score 37; DB 13; Length 234;
Best Local Similarity 75.0%; Pred. No. 48;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 LVNERNWL 9
  ||:|||||
Db 30 LVNERNWL 37

RESULT 14
Q9VEM7 PRELIMINARY; PRT; 243 AA.
AC Q9VEM7
D Q9VEM7
T 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE CG4053 protein.
GN CG4053
OS Drosophila melanogaster (fruit fly).
CC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
CC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
CC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,
RA Wan X.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agapayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,

```

A de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 A Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 A Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 A Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 A Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 A Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 A Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,  
 A Jajali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 A Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 A Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 A Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 A Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 A Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 A Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacלב J.M.,  
 A Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 A Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 A Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 A Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 A Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 A Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
 A Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 A Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 A Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 A Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 T "The genome sequence of *Drosophila melanogaster*.";  
 L Science 287:2185-2185(2000).  
 L C -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.  
 R EMBL; AF003716; AAF5394.1; -.  
 R HSP; P00766; 1CHG.  
 R FlyBase; FBgn0038482; CG4053.  
 R GO; GO:0004263; F:chymotrypsin activity; IEA.  
 R GO; GO:0008233; F:peptidase activity; IEA.  
 R GO; GO:0004295; F:trypsin activity; IEA.  
 R GO; GO:0008508; P:proteolysis and peptidolysis; IEA.  
 R InterPro; IPR009003; Cys Ser trypsin.  
 R InterPro; IPR001254; Peptidase\_S1.  
 R InterPro; IPR001314; Peptidase\_S1A.  
 R Pfam; PF00089; trypsin; 1.  
 R PRINTS; PR00722; CHYMOTRYPSIN.  
 R SMART; SM00020; TRY SPG; 1.  
 R PROSITE; PS0240; TRYPSIN\_DOM; 1.  
 R PROSITE; PS00134; TRYPSIN\_HIS; 1.  
 R PROSITE; PS00135; TRYPSIN\_SER; 1.  
 W Hydrolase; Protease; Serine protease.  
 Q SEQUENCE 243 AA; 2692 MW; 393C4B710563C811 CRC64;

Query Match 78.7%; Score 37; DB 5; Length 243;  
 Best Local Similarity 55.6%; Pred. No. 50;  
 Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

1 VLVNERWVL 9  
 ::::|::|  
 65 VILNEQWIL 73

RESULT 15  
 19W6K0

D Q9W6K0 PRELIMINARY; PRT; 249 AA.  
 C Q9W6K0;  
 T 01-NOV-1999 (TEMBLrel. 12, Created)  
 T 01-NOV-1999 (TEMBLrel. 12, Last sequence update)  
 T 01-OCT-2003 (TEMBLrel. 25, Last annotation update)  
 E Trypsinogen-like serine protease.  
 S Notothenia coriiceps (black rockcod).  
 C Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 C Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 C Acanthomorpha; Acanthopterygii; Percormorpha; Perciformes;  
 C Nototheniidae; Nototheniidae; Notothenia.  
 X NCBI\_TaxID=8208;  
 X [1]

SEQUENCE FROM N.A.  
 Cheng C.-H.C., Chen L.;  
 "Evolutionary transition from serine protease to antifreeze

RT glycoprotein captured.";  
 RL Nature 0:0-0(1999).  
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.  
 DR EMBL; AF134323; AAD30107.1; -.  
 DR HSP; P00761; 1EFT.  
 DR MEROPS; S01.134; -.  
 DR GO; GO:0004263; F:chymotrypsin activity; IEA.  
 DR GO; GO:0008233; F:peptidase activity; IEA.  
 DR GO; GO:0004295; F:trypsin activity; IEA.  
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR InterPro; IPR009003; Cys Ser trypsin.  
 DR InterPro; IPR001354; Peptidase\_S1.  
 DR InterPro; IPR001314; Peptidase\_S1A.  
 DR PRINTS; PR00722; CHYMOTRYPSIN.  
 DR Pfam; PF00089; trypsin; 1.  
 DR SMART; SM00020; TRY SPG; 1.  
 DR PROSITE; PS0240; TRYPSIN\_DOM; 1.  
 DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
 KW Hydrolase; Protease; Serine protease.  
 SQ SEQUENCE 249 AA; 27419 MW; 69C0D872DFB96ED6 CRC64;

Query Match 78.7%; Score 37; DB 13; Length 249;  
 Best Local Similarity 56.7%; Pred. No. 51;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 VLVNERWVL 9  
 ::::|::|  
 DB 50 VILNEQWVL 58

Search completed: March 1, 2004, 17:34:47  
 Job time : 33.3333 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

M protein - protein search, using sw model

un on: March 1, 2004, 17:17:25 ; Search time 6 Seconds  
(without alignments)  
78.105 Million cell updates/sec

title: US-09-905-083-34

effect score: 47

sequence: 1 VLVNERWYL 9

scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

sarched: 141681 seqs, 52070155 residues

total number of hits satisfying chosen parameters: 141681

inimum DB seq length: 0

aximum DB seq length: 2000000000

ost-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

atabase : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

result No.	Score	Query Match	Length	ID	Description
1	47	100.0	253	1 KLK7_HUMAN	P49662 homo sapien
2	38	80.9	246	1 KLK_PIG	P00752 sus scrofa
3	38	80.9	260	1 NRPN_MOUSE	O61955 mus musculus
4	38	80.9	260	1 NRPN_RAT	O88780 rattus norv
5	37	78.7	250	1 TRYP_PLEPL	P35034 pleuronecte
6	37	78.7	257	1 VSP4_TRIMU	O91510 trimeresuru
7	37	78.7	257	1 VSP5_TRIMU	O91511 trimeresuru
8	37	78.7	257	1 VSP7_TRIMU	O90984 trimeresuru
9	37	78.7	257	1 VSPC_TRIGA	O13062 trimeresuru
10	37	78.7	258	1 CTR2_ANOGA	Q17025 anopheles g
11	37	78.7	259	1 CTR1_ANOGA	Q27289 anopheles g
12	37	78.7	260	1 VSP1_TRIJE	O94f68 trimeresuru
13	37	78.7	260	1 VSP2_TRIJE	O13057 trimeresuru
14	36	76.6	235	1 ATE_PSEPK	O88f56 pseudomonas
15	36	76.6	248	1 KLRK_HUMAN	O90k00 homo sapien
16	36	76.6	258	1 GRAM_RAT	O30238 rattus norv
17	36	76.6	260	1 ESTA_CANFA	P09582 canis famil
18	36	76.6	345	1 TAR4_RAT	Q923Y5 rattus norv
19	36	76.6	426	1 VA20_VACCC	P20995 vaccinia vi
20	36	76.6	426	1 VA20_VARV	P33843 variola vir
21	36	76.6	516	1 X4BL_RHISN	P55379 rhizobium s
22	35	74.5	231	1 TRIZ_SALSA	P35032 salmo salar
23	35	74.5	242	1 TRV1_SALSA	P35031 salmo salar
24	35	74.5	247	1 TRV1_HUMAN	P07477 homo sapien
25	35	74.5	253	1 CFAD_HUMAN	P00746 homo sapien
26	35	74.5	257	1 VSP1_TRIMU	O91507 trimeresuru
27	35	74.5	257	1 VSP2_TRIMU	O91508 trimeresuru
28	35	74.5	257	1 VSP3_AKAC	O918x0 agkistrodon
29	35	74.5	257	1 VSP3_TRIMU	O91509 trimeresuru
30	35	74.5	274	1 FA9_SHEEP	P16291 ovis aries
31	35	74.5	274	1 RS2_AQUAE	O67809 aquifex aeo
32	35	74.5	324	1 TEST_MOUSE	O91hj7 mus musculu
33	35	74.5	329	1 HPT_CANFA	P19006 canis famil

34	35	74.5	347	1 HPT_ATEGE	P50417 ateles geof
35	35	74.5	347	1 HPT_PIG	O88p67 sus scrofa
36	35	74.5	348	1 HPT_HUMAN	P00739 homo sapien
37	35	74.5	406	1 HPT_HUMAN	P00738 homo sapien
38	35	74.5	413	1 ECA3_ARATH	Q9m401 arabidopsis
39	35	74.5	416	1 FA9_BOVIN	P00741 bos taurus
40	35	74.5	421	1 ACRO_HUMAN	P10323 homo sapien
41	35	74.5	688	1 C1S_HUMAN	P09871 homo sapien
42	34	72.3	178	1 YOAC_BACSU	P45900 bacillus su
43	34	72.3	228	1 VSPA_LACMU	P33589 lachesis mu
44	34	72.3	231	1 VSP1_AKCO	P09872 agkistrodon
45	34	72.3	232	1 VSP1_BOTJA	P81824 bothrops ja

#### ALIGNMENTS

RESULT 1  
 KLK7\_HUMAN  
 ID KLK7\_HUMAN STANDARD; PRT; 253 AA.  
 AC P49862;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Kallikrein 7 precursor (EC 3.4.21.-) (Stratum corneum chymotryptic  
 enzyme) (hSCCE).  
 DE enzyme (hSCCE).  
 GN KLK7 OR PRS86 OR SCCE.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 23-53.  
 RC TISSUE=Skin;  
 RX MEDLINE=94308225; PubMed=8034709;  
 RA Hansson L., Stromqvist M., Baackman A., Wallbrandt P., Carlstein A.,  
 RA Egelrud T.;  
 RT "Cloning, expression, and characterization of stratum corneum  
 RT chymotryptic enzyme. A skin-specific human serine proteinase.";  
 RL J. Biol. Chem. 269:19420-19426(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Keratinocytes;  
 RA Yousef G.M., Scorillas A., Diamandis E.P.;  
 RT "Molecular characterization, mapping and tissue expression of the  
 RT human stratum corneum chymotryptic enzyme gene.";  
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20510030; PubMed=11054574;  
 RA Gan L., Lee I., Smith R., Argonza-Barrett R., Lei H., McCuaig J.,  
 RA Moss P., Paepker B., Wang K.;  
 RT "Sequencing and expression analysis of the serine protease gene  
 RT cluster located in chromosome 19q13 region.";  
 RL Gene 257:119-130(2000).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Hansson L., Backman A., Ny A., Edlund M., Edholm E., Tornell J.,  
 RA Wallbrandt P., Egelrud T.;  
 RT "Epidermal overexpression of stratum corneum chymotryptic enzyme in  
 RT mice: a model for chronic itchy dermatitis.";  
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP CHARACTERIZATION.  
 RX MEDLINE=95314630; PubMed=7794273;  
 RA Skvtt A., Stromqvist M., Egelrud T.;  
 RT "Primary substrate specificity of recombinant human stratum corneum  
 RT chymotryptic enzyme.";  
 RL Biochem Biophys. Res. Commun. 211:586-589(1995).  
 CC -!- FUNCTION: May catalyze the degradation of intercellular cohesive  
 CC structures in the cornified layer of the skin in the continuous  
 CC shedding of cells from the skin surface. Specific for amino acid  
 CC residues with aromatic side chains in the P1 position. SCCE

cleaves insulin B chain at 6-Leu-|-Cys-7, 16-Tyr-|-Leu-17, 25-Phe-|-Tyr-26, and 26-Tyr-|-Thr-27. Could play a role in the activation of precursors to inflammatory cytokines.

-!- TISSUE SPECIFICITY: Abundantly expressed in the skin and is expressed by keratinocytes in the epidermis. Very low levels are also seen in the brain and kidney.

-!- SIMILARITY: Belongs to peptidase family S1. Kallikrein subfamily.

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

EMBL; L33404; AAC37551.1; -  
EMBL; AF166330; AAD49718.1; -  
EMBL; AF243527; AAG33360.1; -  
EMBL; AF332583; AAK59624.1; -  
PIR; A53968; A53968.  
HSSP; P00763; IDPO.  
MEROPS; S01.300; -  
Genew; HGNC:6368; KLK7.  
MIM; 604438; -  
GO; GO:0008236; F:serine-type peptidase activity; TAS.  
GO; GO:0008544; P:epidermal differentiation; TAS.  
InterPro; IPR009003; Cys\_ser\_trypsin.  
InterPro; IPR001254; Peptidase\_S1.  
InterPro; IPR001314; Peptidase\_S1A.  
Pfam; PF00089; trypsin; 1.  
PRINTS; PR00722; CHYMOTRYPSIN.  
SMART; SM00020; Tryp\_SPC; 1.  
PROSITE; PS00240; TRYPSIN\_DOM; 1.  
PROSITE; PS00134; TRYPSIN\_HIS; 1.  
PROSITE; PS00135; TRYPSIN\_SER; 1.  
Hydrolase; Serine protease; Zymogen; Glycoprotein; Signal.  
SIGNAL  
1 22  
PROPEP 23 29  
CHAIN 30 253  
KALLIKREIN 7.  
ACT\_SITE 70 70  
CHARGE RELAY SYSTEM (BY SIMILARITY).  
ACT\_SITE 112 112  
CHARGE RELAY SYSTEM (BY SIMILARITY).  
ACT\_SITE 205 205  
CHARGE RELAY SYSTEM (BY SIMILARITY).  
BY SIMILARITY.  
DISULFID 36 137  
BY SIMILARITY.  
DISULFID 55 71  
BY SIMILARITY.  
DISULFID ? 239  
BY SIMILARITY.  
DISULFID 144 211  
BY SIMILARITY.  
DISULFID 176 190  
BY SIMILARITY.  
DISULFID 201 226  
BY SIMILARITY.  
CARBOHYD 246 246  
N-LINKED (GLCNAC... ) (POTENTIAL).  
SEQUENCE 253 AA; 27525 MW; 2D68B8B18A76A668 CRC64;  
Query Match 100.0%; Score 47; DB 1; Length 253;  
Best Local Similarity 100.0%; Pred.No. 0.082;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Y 1 VLVNRRWVL 9  
b 58 VLVNRRWVL 66  
-----  
RESULT 2  
LK\_PIG  
D\_KLK\_PIG STANDARD; PRT; 246 AA.  
C P00752;  
T 21-JUL-1986 (Rel. 01, Created)  
T 28-FEB-2003 (Rel. 41, Last sequence update)  
T 10-OCT-2003 (Rel. 42, Last annotation update)  
S Glandular kallikrein precursor (EC 3.4.21.35) (Tissue kallikrein).  
E Sus scrofa (Pig).  
C Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
C Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
X NCBI\_TaxId=9823;

-----

SEQUENCE OF 1-15 AND 95-102.  
MEDLINE=89230643; PubMed=3246048;  
RA Kamada M., Aoki K., Ikekita M., Kizuki K., Moriya H., Kamo M.,  
RA Tsugita A.;  
RT "Generation of alpha- and beta-kallikreins from porcine pancreatic  
prokallikrein by the action of trypsin";  
RL Chem. Pharm. Bull. 36:4891-4899(1988).  
[2]  
SEQUENCE OF 8-87 AND 95-246.  
TISSUE=Pancreas;  
RA Tschesche H., Mair G., Godec G., Fiedler F., Ehret W., Hirschauer C.,  
RA Lemon M., Fritz H., Schmidt-Kastner G., Kutzbach C.;  
RT "The primary structure of porcine glandular kallikreins";  
RL Adv. Exp. Med. Biol. 120:245-260(1979).  
[3]  
SEQUENCE OF 8-87; 95-127 AND 176-246.  
TISSUE=Pancreas;  
RA Ehret W.;  
RT "The primary structure of the kallikrein from porcine pancreas";  
RL Thesis (1976), University of Munich, Germany.  
[4]  
SEQUENCE OF 84-98.  
MEDLINE=90336040; PubMed=2379280;  
RA Kamada M., Ikekita M., Kuranashi T., Aoki K., Kizuki K., Moriya H.,  
RA Sweeley C.C., Kamo M., Tsugita A.;  
RT "Generation of a different type of beta-kallikrein from porcine  
pancreatic alpha-kallikrein by the action of chymotrypsin --  
observation of proteolytic processing occurring around 'kallikrein  
autolysis loop' region";  
RL Chem. Pharm. Bull. 38:1053-1057(1990).  
[5]  
SEQUENCE OF 128-175.  
TISSUE=Pancreas;  
RA Ehret W.;  
RT "Investigation of the sequence of amino acid residues 127 to 174 of  
the kallikrein from porcine pancreas";  
RL Thesis (1978), University of Munich, Germany.  
[6]  
MEDLINE=82194877; PubMed=7043199;  
RA Fiedler F., Fink E., Tschesche H., Fritz H.;  
RT "Porcine glandular kallikreins";  
RL Meth. Enzymol. 80:493-532(1981).  
[7]  
X-RAY CRYSTALLOGRAPHY (2 ANGSTROMS), AND REVISIONS.  
RA Bode W., Henz Z., Bartels K., Kutzbach C., Schmidt-Kastner G.,  
RA Bartunik H.;  
RT "Refined 2-A X-ray crystal structure of porcine pancreatic kallikrein  
A, a specific trypsin-like serine proteinase. Crystallization,  
structure determination, crystallographic refinement, structure and  
its comparison with bovine trypsin";  
RL J. Mol. Biol. 164:237-282(1983).  
[8]  
X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF COMPLEX WITH BOVINE  
PANCREATIC TRYPSIN INHIBITOR.  
RA Chen Z., Bode W.;  
RT "Refined 2.5 A X-ray crystal structure of the complex formed by  
porcine kallikrein A and the bovine pancreatic trypsin inhibitor.  
Crystallization, Patterson search, structure determination,  
refinement, structure and comparison with its components and with the  
bovine trypsin-pancreatic trypsin inhibitor complex";  
RL J. Mol. Biol. 164:283-311(1983).  
[9]  
X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF COMPLEX WITH HIRUSTASIN.  
MEDLINE=97184690; PubMed=9032072;  
RA Mattl P.R.E., di Marco S., Fendrich G., Pohlig G., Heim J.,  
RA Sommerhoff C., Fritz H., Priestle J.P., Gruetter M.G.;  
RT "A new structural class of serine protease inhibitors revealed by the  
structure of the hirustasin-kallikrein complex";  
RL Structure 5:253-264(1997).

[10]  
 ERRATUM.  
 Mittl P.R.E., di Marco S., Fendrich G., Pohlig G., Heim J.,  
 Sommerhoff C., Fritz H., Priestle J.P., Gruetter M.G.;  
 Structure 5:585-585(1997).  
 [11]  
 STRUCTURE OF CARBOHYDRATES.  
 MEDLINE=89062455; PubMed=3196708;  
 Tomiya N., Yamaguchi T., Awaya J., Kurono M., Endo S., Arata Y.,  
 Takahashi N., Ishihara H., Mori M., Tejima S.;  
 "Structural analyses of asparagine-linked oligosaccharides of porcine  
 pancreatic kallikrein";  
 Biochemistry 27:7146-7154(1988).  
 -!- FUNCTION: Glandular kallikreins cleave Met-Lys and Arg-Ser bonds  
 in kininogen to release Lys-bradykinin.  
 -!- CATALYTIC ACTIVITY: Preferential cleavage of Arg-|-Xaa bonds in  
 small molecule substrates. Highly selective action to release  
 kallidin (lysyl-bradykinin) from kininogen involves hydrolysis of  
 Met-|-Xaa or Leu-|-Xaa.  
 -!- SUBUNIT: Monomer.  
 -!- SIMILARITY: Belongs to peptidase family S1. Kallikrein subfamily.  
 -!- CAUTION: Native porcine kallikrein is a monomer. Chains of the  
 pancreatic beta-kallikrein are heterogeneous artifacts of  
 proteolytic degradation during isolation.  
 PDB; 2PKA; 19-JUL-84.  
 PDB; 2XAI; 15-OCT-91.  
 PDB; 1HIA; 24-DEC-97.  
 MEROPS; S01.160; --  
 InterPro; IPR009003; Cys Ser trypsin.  
 InterPro; IPR001254; Peptidase\_S1.  
 InterPro; IPR001314; Peptidase\_S1A.  
 Pfam; PF00089; trypsin; 1.  
 PRINTS; PRO0722; CHYMOTRYPSIN.  
 SMART; SM00020; Tryp SPc; 1.  
 PROSITE; PS00240; TRYPSIN\_DOM; 1.  
 PROSITE; PS00134; TRYPSIN\_HIS; 1.  
 PROSITE; PS00135; TRYPSIN\_SER; 1.  
 Hydrolyase; Serine protease; Glycoprotein; Pancreas; 3D-structure;  
 Zymogen.  
 PROPEP  
 CHAIN 1 8 246 7  
 DOMAIN 85 104 158  
 DISULFID 14 158  
 DISULFID 33 49  
 DISULFID 135 204  
 DISULFID 169 183  
 DISULFID 194 219  
 CARBOHYD 85 85  
 CARBOHYD 239 239  
 ACT\_SITE 48 48  
 ACT\_SITE 103 103  
 ACT\_SITE 198 198  
 STRAND 9 9  
 TURN 10 10  
 STRAND 12 13  
 TURN 16 17  
 TURN 20 21  
 STRAND 22 27  
 TURN 28 29  
 STRAND 30 39  
 TURN 40 41  
 STRAND 42 45  
 HELIX 47 49  
 STRAND 55 58  
 STRAND 62 62  
 TURN 63 64  
 STRAND 71 73  
 STRAND 75 80  
 TURN 82 83  
 TURN 97 98  
 STRAND 105 109  
 TURN 131 132  
 STRAND 134 139

FT STRAND 155  
 FT STRAND 157  
 FT HELIX 164  
 FT TURN 169  
 FT TURN 170  
 FT TURN 172  
 FT TURN 179  
 FT STRAND 180  
 FT TURN 181  
 FT TURN 185  
 FT TURN 187  
 FT STRAND 188  
 FT TURN 192  
 FT TURN 195  
 FT TURN 198  
 FT STRAND 201  
 FT TURN 205  
 FT STRAND 207  
 FT TURN 214  
 FT TURN 219  
 FT TURN 220  
 FT TURN 222  
 FT STRAND 223  
 FT TURN 226  
 FT HELIX 230  
 FT STRAND 231  
 SQ SEQUENCE 246 AA; 27170 MW; 5991CEDE406A19A1 CRC64;  
 Query Match 80.9%; Score 38; DB 1; Length 246;  
 Best Local Similarity 77.8%; Pred. No. 4.3;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 VLVNERWVL 9  
 DB 36 VLVNPKWVL 44  
 RESULT 3  
 NRPN MOUSE STANDARD; PRT; 260 AA.  
 ID NRPN MOUSE  
 AC Q61955;  
 DT 15-JUL-1999 (Rel. 38, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Neuropilin precursor (EC 3.4.21.-) (NP) (Kallikrein 8).  
 GN K1K8 OR PRSS19 OR NRPN.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BALB/c; TISSUE=Hippocampus;  
 RX MEDLINE=95348817; PubMed=7623137;  
 RA Chen Z.-I., Yoshida S., Kato K., Momota Y., Suzuki J., Tanaka T.,  
 Ito J., Nishino H.,imoto S., Kiyama H., Shiosaka S.,  
 RT "Expression and activity-dependent changes of a novel limbic-serine  
 RT protease gene in the hippocampus."  
 RL J. Neurosci. 15:5088-5097(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Yoshida S., Hirata A., Inoue N., Shiosaka S.;  
 RT "Cloning and assignment of mouse neuropilin gene, Prss19 to chromosome  
 RT 7B4.";  
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Czech II; TISSUE=Mammary gland;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Schuler G.D.,  
 Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schaefer C.F., Bhat N.K.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max J., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,

Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
Rodriguez A.C., Grimwood J., Schmitt J., Myers R.M.,  
Butterfield Y.S.N., Krzyzinski M.I., Skalska U., Smailus D.E.,  
Schnerch A., Schein J.E., Jones S.J.M., Maria M.A.,  
"Generation and initial analysis of more than 15,000 full-length  
human and mouse cDNA sequences".  
Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
[4]  
SEQUENCE OF N-TERMINUS, AND CHARACTERIZATION.  
STRAIN=BAIB/C; TISSUE=Brain;  
MEDLINE=98225202; PubMed=9556608;  
Shimizu C., Yoshida S., Shibata M., Kato K., Momota Y., Matsumoto K.,  
Shiosaka T., Midorikawa R., Kamachi T., Kawabe A., Shiosaka S.,  
"Characterization of recombinant and brain neuropsin, a  
plasticity-related serine protease".  
J. Biol. Chem. 273:11189-11196 (1998).  
[5]  
X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 33-257.  
TISSUE=Hippocampus;  
MEDLINE=95134351; PubMed=9933620;  
Kishi T., Kato M., Shimizu T., Kato K., Matsumoto K., Yoshida S.,  
Shiosaka S., Hakoshima T.,  
"Crystal structure of neuropsin, a hippocampal protease involved in  
kindling epileptogenesis".  
J. Biol. Chem. 274:4220-4224 (1999).  
-!- FUNCTION: Suggested to be involved in kindling epileptogenesis and  
hippocampal plasticity. Has a strong proteolytic activity against  
fibrinectin.  
-!- CATALYTIC ACTIVITY: Preferential cleavage: Arg-, Lys-.  
-!- ENZYME REGULATION: Strongly inhibited by diisopropyl  
fluorophosphate, leupeptin and (4-aminophenyl)methanesulfonyl 1-  
fluoride.  
-!- SUBCELLULAR LOCATION: Secreted.  
-!- TISSUE SPECIFICITY: Expressed specifically in the limbic system of  
mouse brain and is localized at highest concentration in pyramidal  
neurons of the hippocampal CA1-3 subfields.  
-!- MASS SPECTROMETRY: MW=26613; METHOD=MALDI; RANGE=29-260.  
-!- MASS SPECTROMETRY: MW=26229; METHOD=MALDI; RANGE=33-260.  
-!- SIMILARITY: Belongs to peptidase family S1. Kallikrein subfamily.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

EMBL; D30785; BAA06451.1; --  
EMBL; AB032202; BAA92435.1; --  
EMBL; BC055895; BAA55895.1; --  
FIR; I56559; I56559.  
PDB; INPM; 23-MAR-99.  
MEROPS; S01.244; --  
MGD; MGI:892018; Klk8.  
InterPro; IPR009003; Cys Ser trypsin.  
InterPro; IPR001254; Peptidase\_S1.  
InterPro; IPR001314; Peptidase\_S1A.  
Pfam; PF00089; trypsin; 1.  
PRINTS; PRO0722; CHYMOTRYPSIN.  
SMART; SM00020; Tryp\_SPC; 1.  
PROSITE; PS0240; TRYPsin\_DOM; 1.  
PROSITE; PS00134; TRYPsin\_HIS; 1.  
PROSITE; PS00135; TRYPsin\_SER; 1.  
Hydrolase; Serine protease; Glycoprotein; Zymogen; Signal;  
3D-structure. 1 28 POTENTIAL.  
SIGNAL 29 32  
PROPEP 33 260  
CHAIN 73 73  
ACT\_SITE 120 120  
ACT\_SITE 212 212  
ACT\_SITE 212 212  
DISULFID 39 173

FT DISULFID 58 74  
FT DISULFID 145 246  
FT DISULFID 152 218  
FT DISULFID 184 198  
FT DISULFID 208 233  
FT CARBOHYD 110 110 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT STRAND 34 34  
FT STRAND 37 38  
FT TURN 41 42  
FT TURN 45 46  
FT STRAND 47 52  
FT STRAND 53 54  
FT STRAND 55 64  
FT TURN 65 66  
FT STRAND 67 70  
FT HELIX 72 74  
FT STRAND 80 83  
FT STRAND 87 87  
FT TURN 88 89  
FT STRAND 96 98  
FT STRAND 100 105  
FT TURN 107 108  
FT TURN 114 115  
FT TURN 118 119  
FT STRAND 122 126  
FT STRAND 140 141  
FT TURN 148 149  
FT STRAND 151 156  
FT STRAND 170 170  
FT STRAND 172 178  
FT HELIX 181 187  
FT TURN 189 191  
FT TURN 194 195  
FT STRAND 196 200  
FT TURN 202 203  
FT STRAND 206 206  
FT TURN 209 210  
FT TURN 212 213  
FT STRAND 215 218  
FT TURN 219 220  
FT STRAND 221 228  
FT STRAND 235 235  
FT TURN 236 237  
FT STRAND 238 238  
FT STRAND 240 244  
FT HELIX 245 256  
SQ SEQUENCE 260 AA; 28523 MW; BESF6F6B37CD60E CRC64;  
Query Match 80.9%; Score 38; DB 1; Length 260;  
Best Local Similarity 77.8%; Pred. No. 4.5;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 1 VLVNERWVL 9  
DB 61 VLVGDRWVL 69  
RESULT 4  
ID\_NRPN\_RAT STANDARD; PRT; 260 AA.  
AC O88780;  
DT 15-JUL-1999 (Rel. 38, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Neuropsin precursor (EC 3.4.21.-) (NP) (Kallikrein 8) (Brain serine  
DE protease 1).  
GN KLK8 OR PRSS19 OR NRPN OR BSP1.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.



STRAIN=Fischer; TISSUE=Brain;  
MEDLINE=98389725; PubMed=9722524;  
Davies B.J., Pickard B.S., Steel M., Morris R.G.M., Lathe R.;  
"Serine proteases in rodent hippocampus";  
J. Biol. Chem. 273:23004-23011(1998).  
-!- FUNCTION: Suggested to be involved in kindling epileptogenesis and  
hippocampal plasticity. Has a strong proteolytic activity against  
fibronectin (By similarity).  
-!- CATALYTIC ACTIVITY: Preferential cleavage: Arg-, Lys-.  
-!- TISSUE SPECIFICITY: Restricted to hippocampus.  
-!- SIMILARITY: Belongs to peptidase family S1. Kallikrein subfamily.  
This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
EMBL; AJ005641; CAA06643.1; -  
HSP; O61955; INPN.  
MEROPS; S01.244; -  
InterPro; IPR009003; Cys Ser trypsin.  
InterPro; IPR001254; Peptidase\_S1.  
InterPro; IPR001314; Peptidase\_S1A.  
Pfam; PF00089; trypsin; 1.  
PRINTS; PR00722; CHYMOTRYPSIN.  
SMART; SM00020; Tryp\_SPC; 1.  
PROSITE; PS00240; TRYPSIN\_DOM; 1.  
PROSITE; PS00134; TRYPSIN\_HIS; 1.  
PROSITE; PS00135; TRYPSIN\_SER; 1.  
Hydrolase; Serine protease; Glycoprotein; Zymogen; Signal.  
FT SIGNAL 1 28  
FT PROPEP 29 32  
FT CHAIN 33 260  
FT ACT\_SITE 73 73 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT\_SITE 120 120 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT\_SITE 212 212 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT DISULFID 39 173 BY SIMILARITY.  
FT DISULFID 58 74 BY SIMILARITY.  
FT DISULFID 145 246 BY SIMILARITY.  
FT DISULFID 152 218 BY SIMILARITY.  
FT DISULFID 184 198 BY SIMILARITY.  
FT DISULFID 208 233 BY SIMILARITY.  
FT CARBOHYD 110 110 N-LINKED (GLCNAC...) (POTENTIAL).  
SQ SEQUENCE 260 AA; 28510 MW; 58DF4F0602A0B7F5 CRC64;  
Query Match 80.9%; Score 38; DB 1; Length 260;  
Best Local Similarity 77.8%; Pred. No. 4.5;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
Y 1 VLVNERWVL 9  
|||:||||  
b 61 VLVGDRWVL 69  
RESULT 5  
D TRYP\_PLEPL STANDARD; PRT; 250 AA.  
C P35034;  
T 01-FEB-1994 (Rel. 28, Created)  
T 01-FEB-1994 (Rel. 28, Last sequence update)  
T 28-FEB-2003 (Rel. 41, Last annotation update)  
E Trypsin precursor (EC 3.4.21.4).  
S Pleuronectes platessa (Plaice).  
C Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
C Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
C Acanthomorpha; Acanthopterygii; Percormorpha; Pleuronectiformes;  
C Pleuronectoidae; Pleuronectidae; Pleuronectes.  
X NCBI\_TaxID=8262;  
N [1]

SEQUENCE FROM N.A.  
RP TISSUE=Liver;  
RA Leaver M.J., George S.G.;  
RL Submitted (NOV-1990) to the EMBL/GenBank/DBJ databases.  
CC -!- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-Xaa, Lys-|-Xaa.  
CC -!- SUBCELLULAR LOCATION: Extracellular.  
CC -!- SIMILARITY: Belongs to peptidase family S1.  
This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
EMBL; X56744; CAA40068.1; -  
PIR; S31384; S31384.  
HSP; P00761; LEPT.  
MEROPS; S01.151; -  
InterPro; IPR009003; Cys\_Ser\_trypsin.  
InterPro; IPR001254; Peptidase\_S1.  
InterPro; IPR001314; Peptidase\_S1A.  
Pfam; PF00089; trypsin; 1.  
PRINTS; PR00722; CHYMOTRYPSIN.  
SMART; SM00020; Tryp\_SPC; 1.  
PROSITE; PS00240; TRYPSIN\_DOM; 1.  
PROSITE; PS00134; TRYPSIN\_HIS; FALSE\_NEG.  
PROSITE; PS00135; TRYPSIN\_SER; 1.  
Hydrolase; Serine protease; Zymogen; Signal.  
FT SIGNAL 1 15  
FT PROPEP 16 22  
FT CHAIN 23 250  
FT ACT\_SITE 62 62 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT\_SITE 106 106 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT\_SITE 203 203 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT DISULFID 29 163 BY SIMILARITY.  
FT DISULFID 47 63 BY SIMILARITY.  
FT DISULFID 133 236 BY SIMILARITY.  
FT DISULFID 140 209 BY SIMILARITY.  
FT DISULFID 174 188 BY SIMILARITY.  
FT DISULFID 199 223 BY SIMILARITY.  
FT SITE 197 197 REQUIRED FOR SPECIFICITY (BY SIMILARITY).  
SQ SEQUENCE 250 AA; 27527 MW; 637DE96185C1ABAA CRC64;  
Query Match 78.7%; Score 37; DB 1; Length 250;  
Best Local Similarity 66.7%; Pred. No. 6.7;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 VLVNERWVL 9  
|||:||||  
Db 50 VLINQWVL 58  
RESULT 6  
VSP4\_TRIMU STANDARD; PRT; 257 AA.  
AC Q91510;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Mucrofibrase 4 precursor (EC 3.4.21.-).  
OS Trimeresurus mucrosquamatus (Taiwan habu).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;  
OC Viperidae; Crotalinae; Protobothrops.  
OX NCBI\_TaxID=103944;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Venom gland;  
RX MEDLINE=95110313; PubMed=7811255;  
RA Hung C.-C., Huang K.F., Chiou S.-H.;  
RT "Characterization of one novel venom protease with beta-fibrinogenase

RT activity from the Taiwan habu (*Trimeresurus mucrosquamatus*):

RT purification and cDNA sequence analysis.;  
 RT Biochem. Biophys. Res. Commun. 205:1707-1715 (1994).  
 CC -!- FUNCTION: Thrombin-like snake venom serine protease. Cleaves beta-chain of fibrinogen molecules efficiently and shows relatively lower activity on alpha-chain, with almost no activity on gamma-chain.  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- TISSUE SPECIFICITY: Expressed by the venom gland.  
 CC -!- SIMILARITY: Belongs to peptidase family S1. Snake venom subfamily.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X83224; CAA58224.1; --  
 CC HSSP; P00763; IDPO.  
 CC MEROPS; S01.185; --  
 CC InterPro; IPR009003; Cys Ser trypsin.  
 CC InterPro; IPR001254; Peptidase\_S1.  
 CC InterPro; IPR001314; Peptidase\_S1A.  
 CC Pfam; PF00089; trypsin; 1.  
 CC PRINTS; PR00722; CHYMOTRYPSIN.  
 CC SMART; SM00020; TRYPSIN\_DOM; 1.  
 CC PROSITE; PS02440; TRYPSIN\_HIS; 1.  
 CC PROSITE; PS00134; TRYPSIN\_HIS; 1.  
 CC PROSITE; PS00135; TRYPSIN\_SER; 1.  
 CC Hydrolase; Serine protease; Zymogen; Signal.

CC SIGNAL 1 18 BY SIMILARITY.  
 CC PROPEP 19 24 BY SIMILARITY.  
 CC CHAIN 25 257 MUCROFIBRASE 4.  
 CC ACT\_SITE 64 64 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 CC ACT\_SITE 109 109 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 CC ACT\_SITE 203 203 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 CC DISULFID 31 162 BY SIMILARITY.  
 CC DISULFID 49 85 BY SIMILARITY.  
 CC DISULFID 97 255 BY SIMILARITY.  
 CC DISULFID 141 209 BY SIMILARITY.  
 CC DISULFID 173 188 BY SIMILARITY.  
 CC DISULFID 199 224 BY SIMILARITY.  
 CC SEQUENCE 257 AA; 28284 MW; 86060B68BCE6FCA8 CRC64;

Query Match 78.7%; Score 37; DB 1; Length 257;  
 Best Local Similarity 75.0%; Pred. No. 6.9;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Y 2 LVNERWVL 9  
 |.|||  
 b 53 LINEEWVL 60

RESULT 7  
 SP5 TRIMU  
 D VSP5 TRIMU STANDARD; PRT; 257 AA.  
 C Q91511;  
 T 16-OCT-2001 (Rel. 40, Created)  
 T 16-OCT-2001 (Rel. 40, Last sequence update)  
 T 28-FEB-2003 (Rel. 41, Last annotation update)  
 E Mucrofibrase 5 precursor (EC 3.4.21.-).  
 S Trimeresurus mucrosquamatus (Taiwan habu).  
 C Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 C Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;  
 C Viperidae; Crotalinae; Protobothrops.  
 X NCBI\_TaxID=103944;  
 X [1]  
 N SEQUENCE FROM N.A.  
 P TISSUE=Venom gland;  
 X MEDLINE=95110313; PubMed=7811255;  
 X Hung C.-C., Huang K.F., Chiou S.-H.;

RT activity from the Taiwan habu (*Trimeresurus mucrosquamatus*):  
 RT purification and cDNA sequence analysis.;  
 RT Biochem. Biophys. Res. Commun. 205:1707-1715 (1994).  
 CC -!- FUNCTION: Thrombin-like snake venom serine protease. Cleaves beta-chain of fibrinogen molecules efficiently and shows relatively lower activity on alpha-chain, with almost no activity on gamma-chain.  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- TISSUE SPECIFICITY: Expressed by the venom gland.  
 CC -!- SIMILARITY: Belongs to peptidase family S1. Snake venom subfamily.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X83225; CAA58225.1; --  
 CC HSSP; P00763; IDPO.  
 CC MEROPS; S01.185; --  
 CC InterPro; IPR009003; Cys Ser trypsin.  
 CC InterPro; IPR001254; Peptidase\_S1.  
 CC InterPro; IPR001314; Peptidase\_S1A.  
 CC Pfam; PF00089; trypsin; 1.  
 CC PRINTS; PR00722; CHYMOTRYPSIN.  
 CC SMART; SM00020; TRYPSIN\_DOM; 1.  
 CC PROSITE; PS02440; TRYPSIN\_HIS; 1.  
 CC PROSITE; PS00134; TRYPSIN\_HIS; 1.  
 CC PROSITE; PS00135; TRYPSIN\_SER; 1.  
 CC Hydrolase; Serine protease; Zymogen; Signal.

CC SIGNAL 1 18 BY SIMILARITY.  
 CC PROPEP 19 24 BY SIMILARITY.  
 CC CHAIN 25 257 MUCROFIBRASE 5.  
 CC ACT\_SITE 64 64 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 CC ACT\_SITE 109 109 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 CC ACT\_SITE 203 203 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 CC DISULFID 31 162 BY SIMILARITY.  
 CC DISULFID 49 85 BY SIMILARITY.  
 CC DISULFID 97 255 BY SIMILARITY.  
 CC DISULFID 141 209 BY SIMILARITY.  
 CC DISULFID 173 188 BY SIMILARITY.  
 CC DISULFID 199 224 BY SIMILARITY.  
 CC SEQUENCE 257 AA; 28164 MW; FCCF64A8EAF2827D CRC64;

Query Match 78.7%; Score 37; DB 1; Length 257;  
 Best Local Similarity 75.0%; Pred. No. 6.9;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 LVNERWVL 9  
 |.|||  
 Db 53 LINEEWVL 60

RESULT 8  
 VSP7 TRIMU  
 ID VSP7 TRIMU STANDARD; PRT; 257 AA.  
 AC Q9DG84;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Serpentokallikrein-2 precursor (EC 3.4.21.-).  
 OS Trimeresurus mucrosquamatus (Taiwan habu).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;  
 OC Viperidae; Crotalinae; Protobothrops.  
 OC NCBI\_TaxID=103944;  
 OC [1]  
 RN SEQUENCE FROM N.A.  
 RP TISSUE=Venom gland;  
 RC Hung C.-C., Huang K.F., Chiou S.-H.;

A Chiu S.-H., Hung C.-C.;  
 P "Serpentokallikrins from Taiwan habu.";  
 L Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
 C -!- FUNCTION: Thrombin-like snake venom serine protease.  
 C -!- SUBCELLULAR LOCATION: Secreted.  
 C -!- TISSUE SPECIFICITY: Expressed by the venom gland.  
 C -!- SIMILARITY: Belongs to peptidase family S1. Snake venom subfamily.  
 C  
 C This SWISS-PROT entry is copyright. It is produced through a collaboration  
 C between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 C the European Bioinformatics Institute. There are no restrictions on its  
 C use by non-profit institutions as long as its content is in no way  
 C modified and this statement is not removed. Usage by and for commercial  
 C entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 C or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 C  
 C EMBL; AF098262; AAG27253.1; -  
 C HSP; P00763; IDPO.  
 C InterPro; IPR009003; Cys\_Ser\_trypsin.  
 C InterPro; IPR001254; Peptidase\_S1.  
 C InterPro; IPR001314; Peptidase\_S1A.  
 C Pfam; PF00089; trypsin; 1.  
 C PRINTS; P00722; CHYMOTRYPSIN.  
 C SMART; SM0020; Tryp\_SPC; 1.  
 C PROSITE; PS00240; TRYPSIN\_DOM; 1.  
 C PROSITE; PS00134; TRYPSIN\_HIS; 1.  
 C PROSITE; PS00135; TRYPSIN\_SER; 1.  
 C W Hydrolase; Serine protease; Glycoprotein; Zymogen; Signal.  
 C T SIGNAL 1 18  
 C T PROPEP 19 24  
 C T CHAIN 25 257  
 C T ACT\_SITE 64 64  
 C T ACT\_SITE 109 109  
 C T ACT\_SITE 203 203  
 C T DISULFID 31 162  
 C T DISULFID 49 65  
 C T DISULFID 97 255  
 C T DISULFID 141 209  
 C T DISULFID 173 188  
 C T DISULFID 199 224  
 C T CARBOHYD 102 102  
 C T SEQUENCE 257 AA; 28320 MW; 61EF5A1DC958F1E9 CRC64;  
 C  
 C Query Match 78.7%; Score 37; DB 1; Length 257;  
 C Best Local Similarity 75.0%; Pred. No. 6.9;  
 C Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 C  
 C Y 2 LVNERWVL 9  
 C :|||  
 C 53 LINEEWVL 60  
 C  
 C RESULT 9  
 C /SPC\_TRIGA STANDARD; PRT; 257 AA.  
 C ID VSFC\_TRIGA  
 C C Q13062; 16-OCT-2001 (Rel. 40, Created)  
 C JT 16-OCT-2001 (Rel. 40, Last sequence update)  
 C JT 16-OCT-2001 (Rel. 40, Last sequence update)  
 C JT 28-FEB-2003 (Rel. 41, Last annotation update)  
 C DE Venom serine proteinase 2C precursor (EC 3.4.21.-).  
 C TN TLG2C.  
 C TS Trimeresurus gramineus (Indian green tree viper) (Green habu snake).  
 C OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 C SC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;  
 C VC Viperidae; Crotalinae; Trimeresurus.  
 C NCBI\_TaxID=8767;  
 C RN [1]  
 C RP SEQUENCE FROM N.A.  
 C RC TISSUE=Venom gland;  
 C RA MEDLINE=97096898; PubMed=8941719;  
 C RA Deshmairu M., Ogawa T., Nakashima K., Nobuhisa I., Chijiwa T.,  
 C RA Shimohigashi Y., Fukumaki Y., Niwa M., Yamashina I., Hattori S.,  
 C RA Ohno M.;

RT "Accelerated evolution of crotalinae snake venom gland serine  
 RT proteases.";  
 RL FEBS Lett. 397:83-88 (1996).  
 CC -!- FUNCTION: Thrombin-like snake venom serine protease.  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- TISSUE SPECIFICITY: Expressed by the venom gland.  
 CC -!- SIMILARITY: Belongs to peptidase family S1. Snake venom subfamily.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; D67084; BAA19982.1; -  
 CC HSP; P00763; IDPO.  
 CC MEROPS; S01.185; -  
 CC InterPro; IPR009003; Cys\_Ser\_trypsin.  
 CC InterPro; IPR001254; Peptidase\_S1.  
 CC InterPro; IPR001314; Peptidase\_S1A.  
 CC Pfam; PF00089; trypsin; 1.  
 CC PRINTS; P00722; CHYMOTRYPSIN.  
 CC SMART; SM0020; Tryp\_SPC; 1.  
 CC PROSITE; PS00240; TRYPSIN\_DOM; 1.  
 CC PROSITE; PS00134; TRYPSIN\_HIS; 1.  
 CC PROSITE; PS00135; TRYPSIN\_SER; 1.  
 CC W Hydrolase; Serine protease; Glycoprotein; Zymogen; Signal.  
 C T SIGNAL 1 18  
 C T PROPEP 19 24  
 C T CHAIN 25 257  
 C T ACT\_SITE 64 64  
 C T ACT\_SITE 109 109  
 C T ACT\_SITE 203 203  
 C T DISULFID 31 162  
 C T DISULFID 49 65  
 C T DISULFID 97 255  
 C T DISULFID 141 209  
 C T DISULFID 173 188  
 C T DISULFID 199 224  
 C T CARBOHYD 116 116  
 C T CARBOHYD 120 120  
 C T CARBOHYD 121 121  
 C T SEQUENCE 257 AA; 28042 MW; 2F50B6947CB33AB1 CRC64;  
 C  
 C Query Match 78.7%; Score 37; DB 1; Length 257;  
 C Best Local Similarity 75.0%; Pred. No. 6.9;  
 C Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 C  
 C QY 2 LVNERWVL 9  
 C :|||  
 C 53 LINEEWVL 60  
 C  
 C RESULT 10  
 C CTF2\_ANOGA STANDARD; PRT; 258 AA.  
 C ID CTR2\_ANOGA  
 C AC Q17025; Q17026; 01-NOV-1997 (Rel. 35, Created)  
 C DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 C DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 C DE Chymotrypsin 2 precursor (EC 3.4.21.1).  
 C GN CHYM2.  
 C OS Anopheles gambiae (African malaria mosquito).  
 C OS Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 C OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anopheles.  
 C NCBI\_TaxID=7165;  
 C RN [1]  
 C RP SEQUENCE FROM N.A.  
 C RC STRAIN=Suakoko; TISSUE=Midgut;  
 C RA Mueller H.M.;  
 C RA Submitted (NOV-1992) to the EMBL/GenBank/DBJ databases.

[2]  
 SEQUENCE FROM N.A.  
 STRAIN=Suakoko; TISSUE=Midgut;  
 Mueller H.M., Catteruccia P., Crisanti A.;  
 "An Anopheles gambiae locus containing the sequences of two closely  
 related chymotrypsin-like proteases induced in the gut following blood  
 meal";  
 Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.  
 Phe-|-Xaa, Leu-|-Xaa.  
 -|- TISSUE SPECIFICITY: Midgut.  
 -|- DEVELOPMENTAL STAGE: Induced in the midgut of female after blood  
 meal.  
 -|- SIMILARITY: Belongs to peptidase family S1.  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 EMBL; Z18888; CAA79326.1; -;  
 EMBL; Z32645; CAA83567.1; -;  
 PIR; S44184; S44184.  
 HSSP; P00761; 1EPT.  
 MEROPS; S01.121; 1.  
 InterPro; IPR009003; Cys\_Ser\_trypsin.  
 InterPro; IPR001254; Peptidase\_S1.  
 InterPro; IPR001314; Peptidase\_S1A.  
 Pfam; PF00089; trypsin; 1.  
 PRINTS; P00722; CHYMOTRYPSIN.  
 SMART; SM00020; Tryp\_SPC; 1.  
 PROSITE; PS0240; TRYPSIN\_DOM; 1.  
 PROSITE; PS00134; TRYPSIN\_HIS; 1.  
 PROSITE; PS00135; TRYPSIN\_SER; 1.  
 HydroLase; Serine protease; Signal; Zymogen; Digestion;  
 Multigene family.  
 SIGNAL 1 17  
 PROPEP 18 32  
 CHAIN 33 258  
 ACT\_SITE 74 74  
 ACT\_SITE 119 119  
 ACT\_SITE 212 212  
 DISULFID 59 75  
 DISULFID 182 198  
 DISULFID 208 232  
 SITE 206 206  
 CONFLICT 134 196  
 CONFLICT 194 206  
 SEQUENCE 258 AA; 27919 MW; A41AD20A630BC67B CRC64;  
 Query Match 78.7%; Score 37; DB 1; Length 258;  
 Best Local Similarity 75.0%; Pred. No. 7;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 Y 2 LVNERWVL 9  
 |:|:|:|  
 b 63 LLNDRWVL 70  
 RESULT 11  
 TR1-ANOCA STANDARD; PRT; 259 AA.  
 C Q27289;  
 T 01-NOV-1997 (Rel. 35, Created)  
 T 01-NOV-1997 (Rel. 35, Last sequence update)  
 T 28-FEB-2003 (Rel. 41, Last annotation update)  
 E Chymotrypsin 1 precursor (EC 3.4.21.1).  
 S CHYM1.  
 S Anopheles gambiae (African malaria mosquito).  
 C Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC NCBI\_TaxId=7165;  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=Suakoko; TISSUE=Midgut;  
 RA Mueller H.M., Catteruccia P., Crisanti A.;  
 RL "An Anopheles gambiae locus containing the sequences of two closely  
 related chymotrypsin-like proteases induced in the gut following blood  
 meal";  
 Submitted (NOV-1992) to the EMBL/GenBank/DBJ databases.  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=Suakoko; TISSUE=Midgut;  
 RA Mueller H.M., Catteruccia P., Crisanti A.;  
 RT "An Anopheles gambiae locus containing the sequences of two closely  
 related chymotrypsin-like proteases induced in the gut following blood  
 meal";  
 Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.  
 Phe-|-Xaa, Leu-|-Xaa.  
 -|- CATALYTIC ACTIVITY: Preferential cleavage: Tyr-|-Xaa, Trp-|-Xaa,  
 CC Phe-|-Xaa, Leu-|-Xaa.  
 CC -|- TISSUE SPECIFICITY: Midgut.  
 CC -|- DEVELOPMENTAL STAGE: Induced in the midgut of female after blood  
 meal.  
 CC -|- SIMILARITY: Belongs to peptidase family S1.  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 EMBL; Z18887; CAA79325.1; -;  
 EMBL; Z32645; CAA83568.1; -;  
 PIR; S49129; S49129.  
 HSSP; P00761; 1EPT.  
 MEROPS; S01.121; 1.  
 InterPro; IPR009003; Cys\_Ser\_trypsin.  
 InterPro; IPR001254; Peptidase\_S1.  
 InterPro; IPR001314; Peptidase\_S1A.  
 Pfam; PF00089; trypsin; 1.  
 PRINTS; P00722; CHYMOTRYPSIN.  
 SMART; SM00020; Tryp\_SPC; 1.  
 PROSITE; PS0240; TRYPSIN\_DOM; 1.  
 PROSITE; PS00134; TRYPSIN\_HIS; 1.  
 PROSITE; PS00135; TRYPSIN\_SER; 1.  
 HydroLase; Serine protease; Signal; Zymogen; Digestion;  
 Multigene family.  
 SIGNAL 1 17  
 PROPEP 18 32  
 CHAIN 33 259  
 ACT\_SITE 74 74  
 ACT\_SITE 119 119  
 ACT\_SITE 212 212  
 DISULFID 59 75  
 DISULFID 182 198  
 DISULFID 208 232  
 SITE 206 206  
 CONFLICT 134 196  
 CONFLICT 194 206  
 SEQUENCE 259 AA; 27717 MW; 66A9C0BA9D1E8C88 CRC64;  
 Query Match 78.7%; Score 37; DB 1; Length 259;  
 Best Local Similarity 75.0%; Pred. No. 7;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 Y 2 LVNERWVL 9  
 |:|:|:|  
 b 63 LLNDRWVL 70  
 RESULT 12  
 VSPI-TRIJE STANDARD; PRT; 260 AA.  
 ID VSPI-TRIJE  
 AC Q9DF68;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)

```

I 28-FEB-2003 (Rel. 41, Last annotation update)
E Venom serine proteinase-like protein precursor (SPI).
S Trimeresurus Jerdonii.
C Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
C Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
C Viperidae; Crotalinae; Trimeresurus.
C NCBI_TaxID=135726;
X [1]
N SEQUENCE FROM N.A.
P TISSUE=Venom gland;
A Lu Q.M., Jin Y., Wei J.F., Wang W.Y., Xiong Y.L.;
C "cDNA cloning of serine proteinases from the venom of Trimeresurus
T Jerdonii."
I Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
L -!- FUNCTION: Snake venom serine protease-like protein. Has lost two
C of the three essential catalytic residues and so probably has no
C enzymatic activity.
C -!- SUBCELLULAR LOCATION: Secreted.
C -!- TISSUE SPECIFICITY: Expressed by the venom gland.
C -!- SIMILARITY: Belongs to peptidase family S1. Snake venom subfamily.
C
C This SWISS-PROT entry is copyright. It is produced through a collaboration
C between the Swiss Institute of Bioinformatics and the EMBL outstation -
C the European Bioinformatics Institute. There are no restrictions on its
C use by non-profit institutions as long as its content is in no way
C modified and this statement is not removed. Usage by and for commercial
C entities requires a license agreement (See http://www.isb-sib.ch/announce/
C or send an email to license@isb-sib.ch).
C -----
C EMBL; AF292110; AAG10788.1; -
C HSP; P00763; IDPO.
C InterPro; IPR009003; Cys_Ser_trypsin.
C InterPro; IPR001254; Peptidase_S1.
C InterPro; IPR001314; Peptidase_S1A.
C Pfam; PF00089; trypsin; 1.
C PRINTS; PR00722; CHYMOTRYPSIN.
C SMART; SM00020; Tryp_SPC; 1.
C PROSITE; PS50240; TRYPSIN_DOM; 1.
C PROSITE; PS00134; TRYPSIN_HIS; FALSE NEG.
C Serine protease homolog; Glycoprotein; Zymogen; Signal.
C SIGNAL 1 18
C PROPEP 19 24
C CHAIN 25 260
C DISULFID 31 165
C DISULFID 52 68
C DISULFID 100 258
C DISULFID 144 212
C DISULFID 176 191
C DISULFID 202 227
C CARBOHYD 123 124
C CARBOHYD 253 253
C SEQUENCE 260 AA; 1692283141E12896 CRC64;
Q
Query Match 78.7%; Score 37; DB 1; Length 260;
Best Local Similarity 75.0%; Pred. No. 7;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
NY 2 LVNERWVL 9
DB 56 LINEEWVL 63
RESULT 13
SP2 TRIFL
D VSP2 TRIFL STANDARD; PRT; 260 AA.
C Q13057;
Y 16-OCT-2001 (Rel. 40, Created)
Y 16-OCT-2001 (Rel. 40, Last sequence update)
Y 28-FEB-2003 (Rel. 41, Last annotation update)
E Venom serine proteinase 2 precursor (EC 3.4.21.-).
N TLF2.
X Trimeresurus flavoviridis (Habu).
C Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC Viperidae; Crotalinae; Trimeresurus.
OC NCBI_TaxID=88087;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Venom gland;
RA MEDLINE=97096898; PubMed=8941719;
RA Deshimaru M., Ogawa T., Nakashima K., Nobuhisa I., Chijiwa T.,
RA Shimohigashi Y., Fukumaki Y., Niwa M., Yamashina I., Hattori S.,
RA Ohno M.;
RA "Accelerated evolution of crotalinae snake venom gland serine
RT proteases." 397.83-88(1996).
RL FEBS Lett. 397.83-88(1996).
CC -!- FUNCTION: Thrombin-like snake venom serine protease.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- TISSUE SPECIFICITY: Expressed by the venom gland.
CC -!- SIMILARITY: Belongs to peptidase family S1. Snake venom subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
C EMBL; D67079; BAA19977.1; -
C HSP; P00763; IDPO.
C MEROPS; S01.185; -
C InterPro; IPR009003; Cys_Ser_trypsin.
C InterPro; IPR001254; Peptidase_S1.
C InterPro; IPR001314; Peptidase_S1A.
C Pfam; PF00089; trypsin; 1.
C PRINTS; PR00722; CHYMOTRYPSIN.
C SMART; SM00020; Tryp_SPC; 1.
C PROSITE; PS50240; TRYPSIN_DOM; 1.
C PROSITE; PS00134; TRYPSIN_HIS; FALSE NEG.
C PROSITE; PS00135; TRYPSIN_SER; FALSE NEG.
C Serine protease homolog; Glycoprotein; Zymogen; Signal.
C SIGNAL 1 18
C PROPEP 19 24
C CHAIN 25 260
C DISULFID 31 165
C DISULFID 52 68
C DISULFID 100 258
C DISULFID 144 212
C DISULFID 176 191
C DISULFID 202 227
C CARBOHYD 123 123
C CARBOHYD 124 124
C SEQUENCE 260 AA; 28641 MW; D0D0A5394CBF9B4A CRC64;
SQ
Query Match 78.7%; Score 37; DB 1; Length 260;
Best Local Similarity 55.6%; Pred. No. 7;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
OY 1 VLVNERWVL 9
DB 55 ILINQEWVL 63
RESULT 14
ATE_PSEPK
ID ATE_PSEPK STANDARD; PRT; 235 AA.
AC Q89FS6;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Probable arginyl-tRNA--protein transferase (EC 2.3.2.8) (R-
DE transferase) (Arginyltransferase).

```

N ATE OR PP4006.  
 S Pseudomonas putida (strain KT2440).  
 C Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
 C Pseudomonadaceae; Pseudomonas.  
 V NCBI\_TaxID=160488;  
 K [1]  
 P SEQUENCE FROM N.A. (ISOFORM 1)  
 K MEDLINE=22423060; PubMed=12534463;  
 A Nelson K.E., Weinel C., Paulsen I.T., Dodson R.J., Hilbert H.,  
 A Martins dos Santos V.A.P., Fouts D.E., Gill S.R., Pop M., Holmes M.,  
 A Brinkac L., Beanan M., DeBoy R.T., Daugherty S., Kolonay J.,  
 A Madupu R., Nelson W., White O., Peterson J., Khouri H., Hance I.,  
 A Chris Lee P., Holtzapple E., Scanlan D., Tran K., Moazzaz A.,  
 A Uterback T., Rizzo M., Lee K., Kosack D., Moestl D., Wedler H.,  
 A Lauber J., Stjepandic D., Hohnsels J., Straetz M., Heim S.,  
 A Kiewitz C., Eisen J.A., Timmis K.N., Duesterhoft A., Thummel B.,  
 A Frazer C.M.;  
 T "Complete genome sequence and comparative analysis of the  
 T metabolically versatile *Pseudomonas putida* KT2440.";  
 L Environ. Microbiol. 4:799-808(2002).  
 C -!- FUNCTION: Functions in the N-end rule pathway of protein  
 C degradation where it conjugates Arg from its aminoacyl-tRNA to the  
 C N-termini of proteins containing an N-terminal aspartate or  
 C glutamate (Potential).  
 C -!- CATALYTIC ACTIVITY: L-arginyl-tRNA + protein = tRNA + L-arginyl-  
 C protein.  
 C -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).  
 C -!- SIMILARITY: Belongs to the R-transferase family.  
 C This SWISS-PROT entry is copyright. It is produced through a collaboration  
 C between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 C the European Bioinformatics Institute. There are no restrictions on its  
 C use by non-profit institutions as long as its content is in no way  
 C modified and this statement is not removed. Usage by and for commercial  
 C entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 C or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 C  
 C EMBL; AE016788; AAN69600.1; -;  
 R TIGR; PP4006; -;  
 R HAMAP; MF\_00689; -; 1.  
 R InterPro; IPR007472; ATE C.  
 R Pfam; PF04377; ATE C; 1.  
 R Pfam; PF04376; ATE N; 1.  
 R Transferase; Acyltransferase; Complete proteome.  
 Q SEQUENCE 235 AA; 27889 MW; F9C30885BCDAE860 CRC64;  
  
 Query Match 76.6%; Score 36; DB 1; Length 235;  
 Best Local Similarity 62.5%; Pred. No. 9.8;  
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;  
  
 Y 1 VLVNERWV 8  
 :||:|  
 b 225 LLINQWV 232  
  
 RESULT 15  
 LKC HUMAN STANDARD; PRT; 248 AA.  
 D Q9UKR0; Q9UKR1;  
 T 16-OCT-2001 (Rel. 40, Created)  
 T 16-OCT-2001 (Rel. 40, Last sequence update)  
 T 16-OCT-2001 (Rel. 40, Last annotation update)  
 E E Kallikrein 12 precursor (EC 3.4.21.-) (Kallikrein-like protein 5)  
 E (KLK-L5).  
 N KLK12 OR KLK15.  
 S Homo sapiens (Human).  
 C Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 C Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 X NCBI\_TaxID=9606;  
 X [1]  
 P SEQUENCE FROM N.A. (ISOFORM 1).  
 X MEDLINE=20118156; PubMed=10652563;

RA Yousef G.M., Luo L.-Y., Diamandis E.P.;  
 RT "Identification of novel human kallikrein-like genes on chromosome  
 RT 19q13.3-q13.4";  
 RL Anticancer Res. 19:2843-2852(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).  
 RA Yousef G.M., Magklara A., Scorilas A., Diamandis E.P.;  
 RT "Cloning of new alternatively spliced forms of the kallikrein-like  
 RT gene 5 (KLK-L5).";  
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RX MEDLINE=20510030; PubMed=11054574;  
 RA Gan L., Lee I., Smith R., Argonza-Barrett R., Lei H., McCuaig J.,  
 RA Moss P., Pieper B., Wang K.;  
 RT "Sequencing and expression analysis of the serine protease gene  
 RT cluster located in chromosome 19q13 region.";  
 RL Gene 257:119-130(2000).  
 RN [4]  
 RP SEQUENCE FROM N.A. (ISOFORM 2).  
 RA Lamerdin J.E., McCready P.M., Skowronski E., Viswanathan V.,  
 RA Burkhardt-Schultz K., Gordon L., Dias J., Ramirez M., Stillwagen S.,  
 RA Phan H., Velasco N., Do L., Regala W., Terry A., Brower A., Gaines J.,  
 RA Dangnan L., Erier A., Christensen M., Georgesdu A., Avila J., Liu S.,  
 RA Andreise T., Frankheim M., Attix C., Amico-Keller G., Cosfield J.,  
 RA Duarte S., Lucas S., Bruce R., Thomas P., Quan G., Kronmiller B.,  
 RA Arellano A., Sanders C., Ow D., Nolan M., Trong S., Kobayashi A.,  
 RA Olsen A.S., Carrano A.V.;  
 RT "Sequence analysis of chromosome 19q13.4";  
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
 CC -!- SUBCELLULAR LOCATION: Secreted (Probable).  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=1;  
 CC IsoId=Q9UKR0-1; Sequences=Displayed;  
 CC Name=2;  
 CC IsoId=Q9UKR0-2; Sequences=VSP 005403;  
 CC -!- SIMILARITY: Belongs to peptidase family S1. Kallikrein subfamily.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AF135025; AAD26426.2; -;  
 DR EMBL; AF135025; AAF06065.1; -;  
 DR EMBL; AF243527; AAG33365.1; -;  
 DR EMBL; AC011473; AAG23258.1; -;  
 DR HSSP; P00763; IDPO.  
 DR MEROPS; S01.020; -;  
 DR Gnew; HGNC:16360; KLK12.  
 DR MIM; 605539; -;  
 DR GO; GO:0005576; C:extracellular; NAS.  
 DR GO; GO:0004252; F:serine-type endopeptidase activity; NAS.  
 DR GO; GO:0006508; P:proteolysis and peptidolysis; NAS.  
 DR InterPro; IPR009003; Cys Ser trypsin.  
 DR InterPro; IPR001254; Peptidase\_S1.  
 DR InterPro; IPR001314; Peptidase\_S1A.  
 DR Pfam; PFO0089; trypsin; 1.  
 DR PRINTS; PR00722; CHYMOTRYPSIN.  
 DR SMART; SMC0020; Tryp SPC; 1.  
 DR PROSITE; PS0240; TRYPsin DOM; 1.  
 DR PROSITE; PS00134; TRYPsin HIS; 1.  
 DR PROSITE; PS00135; TRYPsin SER; 1.  
 DR Hydrolase; Serine protease; Glycoprotein; Signal;  
 XW Alternative splicing.  
 FT SIGNAL 1 17  
 FT CHAIN 18 248  
 FT ACT\_SITE 62 62  
 FT ACT\_SITE 108 108  
 FT ACT\_SITE 108 108  
 FT POTENTIAL.  
 FT KALLIKREIN 12.  
 FT CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT CHARGE RELAY SYSTEM (BY SIMILARITY).

ACT SITE 200 200 CHARGE RELAY SYSTEM (BY SIMILARITY).  
DISULFID 28 161 BY SIMILARITY.  
DISULFID 47 63 BY SIMILARITY.  
DISULFID 133 235 BY SIMILARITY.  
DISULFID 140 206 BY SIMILARITY.  
DISULFID 172 186 BY SIMILARITY.  
DISULFID 196 222 BY SIMILARITY.  
CARBOHYD 24 24 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CARBOHYD 163 163 N-LINKED (GLCNAC. . .) (POTENTIAL).  
VARSPIC 236 248 KYDWMIRMENN -> NSTLVGLGTSWNFNSQPF (in  
isoform 2).  
FTid=VSP.005403.

SEQUENCE 248 AA; 26733 MW; BB473E98F8BAF703 CRC64;

Query Match 76.6%; Score 36; DB 1; Length 248;  
Best Local Similarity 66.7%; Pred. No. 10;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

1 VLVNERWVL 9  
||:||||  
50 VLIHHRWVL 58

search completed: March 1, 2004, 17:29:57  
2b time : 7 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

4 protein - protein search, using sw model

in On: March 1, 2004, 17:16:55 ; Search time 45.5556 Seconds  
(without alignments)  
55.820 Million cell updates/sec

itle: US-09-905-083-35

erfect score: 40

sequence: 1 LLPQLILL 9

oring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

sarched: 1586107 seqs, 282547505 residues

otal number of hits satisfying chosen parameters: 1586107

inimum DB seq length: 0

aximum DB seq length: 2000000000

ost-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

atabase : A\_Geneseq\_29Jan04: \*  
1: Geneseqp1980s: \*  
2: Geneseqp1990s: \*  
3: Geneseqp2000s: \*  
4: Geneseqp2001s: \*  
5: Geneseqp2002s: \*  
6: Geneseqp2003as: \*  
7: Geneseqp2003bs: \*  
8: Geneseqp2004s: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

result No.	Score	Match	Length	DB	ID	Description
1	40	100.0	9	4	AAE08240	Human str
2	40	100.0	136	4	ABG23378	Novel hum
3	40	100.0	198	6	ADA05736	Human NOV
4	40	100.0	250	6	ADA05732	Human NOV
5	40	100.0	253	2	AAR67888	Human str
6	40	100.0	253	2	AAW05383	Human amy
7	40	100.0	253	5	AB884421	Human SCC
8	40	100.0	253	5	AB884406	Human SCC
9	40	100.0	253	5	AAU82740	Amino aci
10	40	100.0	253	6	ABU07440	Protein d
11	40	100.0	253	6	ABU07471	Human str
12	40	100.0	253	6	AB58471	Human d
13	40	100.0	253	7	AD80484	Ovarian c
14	40	100.0	257	3	AA821326	Human HSC
15	38	95.0	142	4	AA863580	Human gas
16	38	95.0	156	4	AA863578	Human gas
17	38	95.0	159	4	AA863582	Human gas
18	36	90.0	9	4	AA808238	Human str
19	36	90.0	845	6	ABU23951	Protein e
20	35	87.5	23	4	AB43858	Peptide #
21	35	87.5	23	4	AA837771	Peptide #
22	35	87.5	23	4	AA864837	Human bra
23	35	87.5	23	4	ABG59233	Human liv
24	35	87.5	23	4	ABG46617	Human pep
25	35	87.5	480	2	AAW98431	H. pylori

26	34	85.0	70	2	AAV36291	Human sec
27	34	85.0	70	6	ADA11692	Human nov
28	34	85.0	73	2	AAW89030	Polypepti
29	34	85.0	73	4	AB851130	Human sec
30	34	85.0	73	6	ABO45387	Novel hum
31	34	85.0	73	7	ABO26867	Protein a
32	34	85.0	86	4	AAO08397	Human pol
33	34	85.0	86	6	ABO08884	Polypepti
34	34	85.0	89	7	ADB09276	Novel pro
35	34	85.0	183	5	AB890096	Human pol
36	34	85.0	201	5	AB878636	Rat OST10
37	34	85.0	224	4	AA893417	Human pro
38	34	85.0	267	2	AAW22303	Rat CRTII
39	34	85.0	301	4	AA887732	Human T2R
40	34	85.0	371	2	AA875642	Bovine co
41	34	85.0	424	2	AAW89026	Polypepti
42	34	85.0	424	4	AB851126	Human sec
43	34	85.0	424	6	ABO45383	Novel hum
44	34	85.0	424	7	ABO26863	Protein a
45	34	85.0	425	6	ABU11726	Human MDD

## ALIGNMENTS

RESULT 1  
AAE08240 standard; peptide; 9 AA.

XX AC AAE08240;  
XX XX  
DT 01-NOV-2001 (first entry)  
XX Human stratum corneum chymotrypsin enzyme peptide #5 (residues 6-14).  
XX Stratum corneum chymotrypsin enzyme; SCCE; cytostatic; vaccine; tumour;  
KW cancer; ovarian; breast; lung; colon; prostate; carcinoma; human;  
KW antisense therapy; malignant hyperplasia.  
XX Homo sapiens.  
XX WO200159158-A1.  
XX 16-AUG-2001.  
XX 07-FEB-2001; 2001WO-US003977.  
XX 11-FEB-2000; 2000US-00502600.  
PR (UYAR-) UNIV ARKANSAS.  
XX O'brien TJ;  
XX WPI; 2001-514676/56.  
XX Diagnosing cancer comprises detecting stratum corneum chymotrypsin enzyme.  
XX Claim 25; Page 103; 127pp; English.

The invention relates to diagnosing cancer especially ovarian cancer, by screening for stratum corneum chymotrypsin enzyme (SCCE). Proteases are considered to be an integral part of tumour growth and metastasis, and therefore, markers indicative of their presence or absence are useful for the diagnosis of cancer. The method is useful for diagnosing cancer, carcinoma and malignant hyperplasia. The SCCE oligonucleotide may be used to treat a cancer selected from ovarian, breast, lung, colon, prostate and other cancers in which SCCE is overexpressed. The present sequence is human SCCE peptide

Sequence 9 AA;

Query Match 100.0%; Score 40; DB 4; Length 9;



Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 LLPLQILL 9  
|||||  
1 LLPLQILL 9

SULT 2  
G23378  
ABG23378 standard; protein; 136 AA.

ABG23378;

18-FEB-2002 (first entry)

Novel human diagnostic protein #23369.

Human; chromosome mapping; gene mapping; gene therapy; forensic;  
food supplement; medical imaging; diagnostic; genetic disorder.

Homo sapiens.

WO200175067-A2.

11-OCT-2001.

30-MAR-2001; 2001WO-US008631.

31-MAR-2000; 2000US-00540217.

23-AUG-2000; 2000US-00649167.

(HYSE-) HYSEQ INC.

Drmanac RT, Liu C, Tang YT;

WPI; 2001-639362/73.

N-PSDB; AAS87565.

New isolated polynucleotide and encoded polypeptides, useful in  
diagnostics, forensics, gene mapping, identification of mutations  
responsible for genetic disorders or other traits and to assess  
biodiversity.

Claim 20; SEQ ID NO 53737; 103pp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II)  
sequences. (I) is useful as hybridisation probes, polymerase chain  
reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
and in recombinant production of (II). The polynucleotides are also used  
in diagnostics as expressed sequence tags for identifying expressed  
genes. (I) is useful in gene therapy techniques to restore normal  
activity of (II) or to treat disease states involving (II). (II) is  
useful for generating antibodies against it, detecting or quantitating a  
polypeptide in tissue, as molecular weight markers and as a food  
supplement. (II) and its binding partners are useful in medical imaging  
of sites expressing (II). (I) and (II) are useful for treating disorders  
involving aberrant protein expression or biological activity. The  
polypeptide and polynucleotide sequences have applications in  
diagnostics, forensics, gene mapping, identification of mutations  
responsible for genetic disorders or other traits to assess biodiversity  
and to produce other types of data and products dependent on DNA and  
amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic  
amino acid sequences of the invention. Note: The sequence data for this  
patent did not appear in the printed specification, but was obtained in  
electronic format directly from WIPO at  
ftp.wipo.int/pub/published\_pct\_sequences

Sequence 136 AA;

Query Match 100.0%; Score 40; DB 4; Length 136;

Best Local Similarity 100.0%; Pred. No. 9.5;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLPLQILL 9  
|||||  
Db 12 LLPLQILL 20

RESULT 3

ADA05736

ID ADA05736 standard; protein; 198 AA.

XX

AC ADA05736;

XX

DT 06-NOV-2003 (first entry)

XX

DE Human NOV18c protein SEQ ID NO:96.

XX

KW human; NOVX; antidiabetic; anorectic; antibacterial; virucide;  
immunomodulator; cytostatic; nootropic; neuroprotective;  
KW antiparkinsonian; antilipemic; gene therapy; human disease;  
KW metabolic disorder; diabetes; obesity; infection; cachexia; cancer;  
KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;  
KW immune disorder; haematopoietic disorder; dyslipidaemia.

XX

OS Homo sapiens.

XX

PN WO2003029424-A2.

XX

PD 10-APR-2003.

XX

PF 02-OCT-2002; 2002WO-US031373.

XX

PR 02-OCT-2001; 2001US-0326483P.

PR 05-OCT-2001; 2001US-0327435P.

PR 09-OCT-2001; 2001US-0327449P.

PR 09-OCT-2001; 2001US-0327917P.

PR 09-OCT-2001; 2001US-0328029P.

PR 09-OCT-2001; 2001US-0328044P.

PR 09-OCT-2001; 2001US-0328056P.

PR 12-OCT-2001; 2001US-0328849P.

PR 15-OCT-2001; 2001US-0329414P.

PR 17-OCT-2001; 2001US-0330142P.

PR 18-OCT-2001; 2001US-0330309P.

PR 22-OCT-2001; 2001US-0341058P.

PR 24-OCT-2001; 2001US-0339266P.

PR 24-OCT-2001; 2001US-0343629P.

PR 24-OCT-2001; 2001US-0349575P.

PR 01-NOV-2001; 2001US-0346357P.

PR 17-APR-2002; 2002US-0373260P.

PR 19-APR-2002; 2002US-0373815P.

PR 19-APR-2002; 2002US-0373817P.

PR 19-APR-2002; 2002US-0373826P.

PR 19-APR-2002; 2002US-0373884P.

PR 22-APR-2002; 2002US-0374977P.

PR 16-MAY-2002; 2002US-0381037P.

PR 16-MAY-2002; 2002US-0381038P.

PR 16-MAY-2002; 2002US-0381042P.

PR 17-MAY-2002; 2002US-0381642P.

PR 28-MAY-2002; 2002US-0383656P.

PR 29-MAY-2002; 2002US-0383831P.

PR 25-JUN-2002; 2002US-0391335P.

PR 01-OCT-2002; 2002US-00262511.

(CURA-) CURAGEN CORP.

XX

XX Smithson G, Millet I, Peyman JA, Kekuda R, Ju J, Guo X;

PI Patturajan M, Spytek KA, Edinger SR, Ellerman K, Malyankar UM;

PI Ort T, Gorman L, Zerhusen BD, Anderson DW, Zhong M, Catterton E;

PI Ji W, Miller CE, Rastelli L, Stone DJ, Pena CBA, Shenoy SG;

PI Shimkets RA, Rotherberg ME, Leach MD, Agse M, Berghs C, Dipippo VA;

PI Eisen AJ, Gangolli EA, Rieger DK, Spaderna SK;

XX WPI; 2003-381626/36.

DR N-PSDB; ADA05735.

X T New NOVX polypeptides and nucleic acids, useful for diagnosing,  
T T preventing or treating NOVX-associated disorders, e.g. diabetes, obesity,  
T T cancer or dyslipidemia, and in chromosome mapping, tissue typing or  
T T pharmacogenomics.  
X S Claim 1; Page 170; 586pp; English.  
X C The present invention describes NOVX proteins, where X can be 1 to 55  
C C (e.g. NOV1). Also described: (1) a composition comprising a polypeptide  
C C described above and a carrier; (2) a kit comprising, in one or more  
C C containers, the composition described above; (3) an isolated nucleic acid  
C C molecule which encodes a NOVX protein of the invention; (4) a vector  
C C comprising the nucleic acid molecule described above; (5) a cell  
C C comprising the above vector; (6) an antibody that immunospecifically  
C C binds to the polypeptide described above; (7) methods for determining the  
C C presence or amount of the above polypeptide or nucleic acid molecule in a  
C C sample; (8) methods for determining the presence of or predisposition to  
C C a disease associated with altered levels of expression of the above  
C C polypeptide or nucleic acid molecule in a first mammalian subject; (9) a  
C C method of identifying an agent that binds to the polypeptide described  
C C above; (10) a method for identifying a potential therapeutic agent for  
C C use in treating a pathology that is related to an aberrant expression or  
C C aberrant physiological interactions of the polypeptide; (11) a method of  
C C screening for a modulator of activity or of latency or predisposition to  
C C a pathology associated with the polypeptide; (12) a method for modulating  
C C the activity of the polypeptide described above; (13) methods of treating  
C C or preventing a pathology associated with the above polypeptide in a  
C C mammal; and (14) a method for producing the above polypeptide. NOVX  
C C sequences have antidiabetic, anorectic, antibacterial, varicidal,  
C C immunomodulator, cytostatic, nootropic, neuroprotective, antiparkinsonian  
C C and antilipidemic activities, and can be used in gene therapy. The  
C C polypeptide is useful in manufacturing a medicament for treating a  
C C syndrome associated with a human disease. The polypeptide or the nucleic  
C C acid molecule may be used to diagnose, treat or prevent metabolic  
C C disorders such as diabetes or obesity, infectious, cachexia, cancer,  
C C neurodegenerative disorders such as Alzheimer's disease or Parkinson's  
C C disease, immune disorders, haematopoietic disorders and various  
C C dyslipidaemias. The nucleic acids can also be used as hybridisation  
C C probes, in chromosome mapping, tissue typing, preventive medicine and  
C C pharmacogenomics. The present sequence represents a human NOVX from the  
C C present invention.  
X X Sequence 198 AA;  
X Q Query Match 100.0%; Score 40; DB 6; Length 198;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Y 1 LLPLQLILL 9  
| | | | |  
Y 6 LLPLQLILL 14  
RESULT 4  
LDA05732  
D ADA05732 standard; protein; 250 AA.  
C ADA05732;  
C ADA05732;  
Y 06-NOV-2003 (first entry)  
X Human NOV18a protein SEQ ID NO:92.  
X human; NOVX; antidiabetic; anorectic; antibacterial; virucide;  
X immunomodulator; cytostatic; nootropic; neuroprotective;  
X antiparkinsonian; antilipidemic; gene therapy; human disease;  
X metabolic disorder; diabetes; obesity; infection; cachexia; cancer;  
X neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;  
X immune disorder; haematopoietic disorder; dyslipidaemia.  
X Homo sapiens.  
X

PN WO2003029424-A2.  
XX 10-APR-2003.  
XX 02-OCT-2002; 2002WO-US031373.  
XX 02-OCT-2001; 2001US-0326483P.  
PR 05-OCT-2001; 2001US-0327435P.  
PR 09-OCT-2001; 2001US-0327449P.  
PR 09-OCT-2001; 2001US-0327917P.  
PR 09-OCT-2001; 2001US-0328029P.  
PR 09-OCT-2001; 2001US-0328044P.  
PR 09-OCT-2001; 2001US-0328056P.  
PR 12-OCT-2001; 2001US-0328849P.  
PR 15-OCT-2001; 2001US-0329414P.  
PR 17-OCT-2001; 2001US-0330142P.  
PR 18-OCT-2001; 2001US-0330309P.  
PR 22-OCT-2001; 2001US-0341058P.  
PR 24-OCT-2001; 2001US-0339266P.  
PR 24-OCT-2001; 2001US-0343629P.  
PR 29-OCT-2001; 2001US-0349575P.  
PR 01-NOV-2001; 2001US-0346357P.  
PR 17-APR-2002; 2002US-0373260P.  
PR 19-APR-2002; 2002US-0373815P.  
PR 19-APR-2002; 2002US-0373817P.  
PR 19-APR-2002; 2002US-0373826P.  
PR 19-APR-2002; 2002US-0373884P.  
PR 22-APR-2002; 2002US-0374977P.  
PR 16-MAY-2002; 2002US-0381037P.  
PR 16-MAY-2002; 2002US-0381038P.  
PR 17-MAY-2002; 2002US-0381642P.  
PR 28-MAY-2002; 2002US-0383656P.  
PR 29-MAY-2002; 2002US-0383831P.  
PR 25-JUN-2002; 2002US-0391335P.  
PR 01-OCT-2002; 2002US-00262511.  
XX (CURA-) CURAGEN CORP.  
XX Smithson G, Millet I, Peyman JA, Kekuda R, Ju J, Li L, Guo X;  
PI Patturajan M, Spytek KA, Edinger SR, Ellerman K, Malyankar UM;  
PI Ort T, Gorman L, Zerhusen BD, Anderson DW, Zhong M, Catterton E;  
PI Ji W, Miller CE, Rastelli L, Stone DJ, Pena CE, Shenoy SG;  
PI Shinkets RA, Rothenberg ME, Leach MD, Agee ML, Berghs C, Dippio VA;  
PI Eisen AJ, Gangolli EA, Rieger DK, Spaderna SK;  
XX WPI; 2003-381626/36.  
DR N-PSDB; ADA05731.  
XX New NOVX polypeptides and nucleic acids, useful for diagnosing,  
PT preventing or treating NOVX-associated disorders, e.g. diabetes, obesity,  
PT cancer or dyslipidemia, and in chromosome mapping, tissue typing or  
PT pharmacogenomics.  
XX Claim 1; Page 169-170; 586pp; English.  
XX The present invention describes NOVX proteins, where X can be 1 to 55  
CC (e.g. NOV1). Also described: (1) a composition comprising a polypeptide  
CC described above and a carrier; (2) a kit comprising, in one or more  
CC containers, the composition described above; (3) an isolated nucleic acid  
CC molecule which encodes a NOVX protein of the invention; (4) a vector  
CC comprising the nucleic acid molecule described above; (5) a cell  
CC comprising the above vector; (6) an antibody that immunospecifically  
CC binds to the polypeptide described above; (7) methods for determining the  
CC presence or amount of the above polypeptide or nucleic acid molecule in a  
CC sample; (8) methods for determining the presence of or predisposition to  
CC a disease associated with altered levels of expression of the above  
CC polypeptide or nucleic acid molecule in a first mammalian subject; (9) a  
CC method of identifying an agent that binds to the polypeptide described  
CC above; (10) a method for identifying a potential therapeutic agent for  
CC use in treating a pathology that is related to an aberrant expression or  
CC aberrant physiological interactions of the polypeptide; (11) a method of  
CC screening for a modulator of activity or of latency or predisposition to

a pathology associated with the polypeptide; (12) a method for modulating the activity of the polypeptide described above; (13) methods of treating or preventing a pathology associated with the above polypeptide in a mammal; and (14) a method for producing the above polypeptide. NOVX sequences have antidiabetic, anorectic, antibacterial, virucide, immunomodulator, cytostatic, nootropic, neuroprotective, antiparkinsonian and anilipaeic activities, and can be used in gene therapy. The polypeptide is useful in manufacturing a medicament for treating a syndrome associated with a human disease. The polypeptide or the nucleic acid molecule may be used to diagnose, treat or prevent metabolic disorders such as diabetes or obesity, infections, cachexia, cancer, neurodegenerative disorders such as Alzheimer's disease or Parkinson's disease, immune disorders, haematopoietic disorders and various dyslipidaemias. The nucleic acids can also be used as hybridisation probes, in chromosome mapping, tissue typing, preventive medicine and pharmacogenomics. The present sequence represents a human NOVX from the present invention.

Sequence 250 AA;

Query Match 100.0%; Score 40; DB 6; Length 250;  
Best Local Similarity 100.0%; Pred. No. 18;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 LLPLQILL 9  
| | | | |  
b 3 LLPLQILL 11

#### RESULT 5

AA067888  
D AAR67888 standard; protein; 253 AA.

AA067888;

25-MAR-2003 (revised)  
09-AUG-1995 (first entry)

Human stratum corneum chymotrophic recombinant enzyme (SCCE).

Stratum corneum chymotryptic enzyme; skin disorder; acne; psoriasis; callosities; keratosis pilaris; ichthyoses; eczema.

Homo sapiens.

WO9500651-Al.

05-JAN-1995.

20-JUN-1994; 94WO-IB000166.

18-JUN-1993; 93DK-00000725.

(SYMB-) SYMBICOM AB.

Egelrud T, Hansson L;

WPI; 1995-052088/07.

N-PSDB; AA081203.

Nucleotide sequences encoding stratum corneum chymotryptic enzyme - and related vectors, transformed cells and polypeptides, useful for treating skin disorders, e.g. acne or psoriasis, and for identification of specific inhibitors.

Disclosure; Page 97; 137pp; English.

The enzyme encoded by this sequence is used in pharmaceutical, cosmetic and skin care products, especially to treat and prevent acne, xeroderma, or other hyperkeratotic conditions (e.g. callosities or keratosis pilaris), ichthyoses, psoriasis, eczema, etc. It is produced recombinantly following mammalian, insect, plant, or microorganism transformation with plasmid pS507. (Updated on 25-MAR-2003 to correct PN

CC field.)

SQ Sequence 253 AA;

Query Match 100.0%; Score 40; DB 2; Length 253;  
Best Local Similarity 100.0%; Pred. No. 18;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLPLQILL 9  
| | | | |  
Db 6 LLPLQILL 14

#### RESULT 6

AAW05383

ID AAW05383 standard; protein; 253 AA.

AAW05383;

31-DEC-1996 (first entry)  
Human amyloid precursor protein protease.

Amyloid precursor protein protease; Alzheimer's disease; diagnosis; therapy.

Homo sapiens.

WO9631122-Al.

10-OCT-1996.

02-APR-1996; 96WO-US004294.

04-APR-1995; 95US-00416257.

(ELIL) LILLY & CO ELI.

Dixon EP, Johnstone EM, Little SP;

WPI; 1996-464694/46.

N-PSDB; AAT39783.

New isolated human amyloid precursor protein protease - used to develop prods. for the treatment or diagnosis of associated conditions, esp. Alzheimer's disease.

Claim 1; Page 44-45; 55pp; English.

Human amyloid precursor protein protease (AAW05383) is involved in the processing or clearance of amyloid precursor protein to form beta-amyloid peptide. Its amino acid sequence was deduced from a cDNA clone (AAT39783) obtid. from a human lung library. Recombinant protease can be produced in transformed or transfected prokaryotic (partic. E. coli) or eukaryotic (partic. AV-120 host cells. It is used to develop products for the design and testing of cpds. useful for treating or preventing conditions associated with beta-amyloid peptide, esp. Alzheimer's disease

Sequence 253 AA;

Query Match 100.0%; Score 40; DB 2; Length 253;  
Best Local Similarity 100.0%; Pred. No. 18;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLPLQILL 9  
| | | | |  
Db 6 LLPLQILL 14

#### RESULT 7

ABB84421

ID ABB84421 standard; peptide; 253 AA.

XX

ABB84421;  
 08-NOV-2002 (first entry)  
 Human SCCE protein N-terminal fragment SEQ ID 48.  
 SCCE; human; stratum corneum chymotryptic enzyme; kallikrein 7;  
 serine protease; transgenic mammal; skin; skin disease; skin cancer;  
 hyperkeratosis; acanthosis; epidermal inflammation; dermal inflammation;  
 pruritus; atopic dermatitis; eczema; acne; itch; KLK7; enzyme.  
 Homo sapiens.  
 WO2000262135-A2.  
 15-AUG-2002.  
 08-FEB-2002; 2002WO-IB001300.  
 09-FEB-2001; 2001CA-02332655.  
 09-FEB-2001; 2001DK-00000218.  
 (EGEL/) EGELRUD T.  
 (HANS/) HANSSON L.  
 Egelrud T, Hansson L;  
 WPI; 2002-643380/59.  
 Transgenic mammal or its embryo useful as model for human disease, has  
 heterologous nucleotide sequence coding for stratum corneum chymotryptic  
 enzyme operably linked to promoter that drives its expression in skin.  
 Example 6; Page 37; 74pp; English.  
 This invention describes a novel non-human transgenic mammal or mammalian  
 embryo having integrated within its genome, a heterologous nucleotide  
 sequence comprising at least a significant part of a nucleotide sequence  
 coding for a stratum corneum chymotryptic enzyme (SCCE) or its variant,  
 operably linked to a promoter that drives expression of heterologous scce  
 or its variant in skin. The product of the invention is useful as a model  
 for the study of disease with the aim of improving treatment, to relieve  
 or ameliorate a pathogenic condition, for development or testing of a  
 cosmetic or a pharmaceutical formulation, and for the development of a  
 diagnostic method. It can also be used as a model for a skin disease or  
 skin cancer. The invention is also useful for screening or identifying a  
 compound or composition effective for the prevention or treatment of an  
 abnormal or unwanted phenotype, and for screening or identifying a  
 compound or composition effective for the prevention or treatment of  
 inflammatory skin diseases selected from diseases consisting of epidermal  
 hyperkeratosis, acanthosis, epidermal inflammation, dermal inflammation,  
 pruritus, atopic dermatitis, eczema, acne and inherited skin diseases  
 with epidermal hyperkeratosis. The mammal of the invention is also useful  
 as a model for further studies of itch mechanisms and the testing of  
 potential compounds and compositions for relieve of various skin diseases  
 where itch is a component. This sequence represents the N-terminal  
 fragment of the human stratum corneum chymotryptic enzyme, SCCE  
 synonymous with human kallikrein 7 (KLK7), used in the development of the  
 transgenic mammals described in the invention  
 Sequence 253 AA;  
 Query Match 100.0%; Score 40; DB 5; Length 253;  
 Best Local Similarity 100.0%; Pred. No. 18;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Y 1 LLPLQILL 9  
 |||||  
 b 6 LLPLQILL 14  
 |||||  
 RESULT 8  
 BB84406

ID XX ABB84406 standard; protein; 253 AA.  
 AC XX ABB84406;  
 XX  
 DT 08-NOV-2002 (first entry)  
 XX  
 DE Human SCCE protein.  
 XX  
 KW SCCE; human; stratum corneum chymotryptic enzyme; kallikrein 7;  
 KW serine protease; transgenic mammal; skin; skin disease; skin cancer;  
 KW hyperkeratosis; acanthosis; epidermal inflammation; dermal inflammation;  
 KW pruritus; atopic dermatitis; eczema; acne; itch; KLK7; enzyme.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO2000262135-A2.  
 XX  
 PD 15-AUG-2002.  
 XX  
 PF 08-FEB-2002; 2002WO-IB001300.  
 XX  
 PP 09-FEB-2001; 2001CA-02332655.  
 PR  
 PR 09-FEB-2001; 2001DK-00000218.  
 XX  
 XX (EGEL/) EGELRUD T.  
 PA (HANS/) HANSSON L.  
 XX  
 XX Egelrud T, Hansson L;  
 PI  
 PI WPI; 2002-643380/59.  
 XX  
 DR N-PSDB; ABQ76226.  
 DR  
 DR Transgenic mammal or its embryo useful as model for human disease, has  
 PT heterologous nucleotide sequence coding for stratum corneum chymotryptic  
 PT enzyme operably linked to promoter that drives its expression in skin.  
 PT  
 XX Claim 10; Page 58-59; 74pp; English.  
 PS  
 XX This invention describes a novel non-human transgenic mammal or mammalian  
 CC embryo having integrated within its genome, a heterologous nucleotide  
 CC sequence comprising at least a significant part of a nucleotide sequence  
 CC coding for a stratum corneum chymotryptic enzyme (SCCE) or its variant,  
 CC operably linked to a promoter that drives expression of heterologous scce  
 CC or its variant in skin. The product of the invention is useful as a model  
 CC for the study of disease with the aim of improving treatment, to relieve  
 CC or ameliorate a pathogenic condition, for development or testing of a  
 CC cosmetic or a pharmaceutical formulation, and for the development of a  
 CC diagnostic method. It can also be used as a model for a skin disease or  
 CC skin cancer. The invention is also useful for screening or identifying a  
 CC compound or composition effective for the prevention or treatment of an  
 CC abnormal or unwanted phenotype, and for screening or identifying a  
 CC compound or composition effective for the prevention or treatment of  
 CC inflammatory skin diseases selected from diseases consisting of epidermal  
 CC hyperkeratosis, acanthosis, epidermal inflammation, dermal inflammation,  
 CC pruritus, atopic dermatitis, eczema, acne and inherited skin diseases  
 CC with epidermal hyperkeratosis. The mammal of the invention is also useful  
 CC as a model for further studies of itch mechanisms and the testing of  
 CC potential compounds and compositions for relieve of various skin diseases  
 CC where itch is a component. This sequence represents the human stratum  
 CC corneum chymotryptic enzyme, SCCE which is a serine protease synonymous  
 CC with human kallikrein 7 (KLK7) and is used in the development of the  
 CC transgenic mammals described in the invention  
 XX  
 XX Sequence 253 AA;  
 SQ  
 Query Match 100.0%; Score 40; DB 5; Length 253;  
 Best Local Similarity 100.0%; Pred. No. 18;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 1 LLPLQILL 9  
 |||||  
 Db 6 LLPLQILL 14  
 |||||

```

> RESULT 9
> U82740
> AAU82740 standard; protein; 253 AA.
>
> AAU82740;
>
> 23-APR-2002 (first entry)
> Amino acid sequence of novel human protease #39.
>
> Human; protease; cancer; immune-related disorder; cardiovascular disease;
> neuronal-associated disease; metabolic disorder; inflammatory disorder;
> nervous system disorder; sexual dysfunction; pain; mood disorder;
> hyperreflexia; psychotic disorder; neurological disorder; dyskinesia;
> viral infection; human immunodeficiency virus; HIV; non-viral infection;
> ocular disease; cytostatic; enzyme.
>
> Homo sapiens.
> WO200200860-A2.
>
> 03-JAN-2002.
>
> 26-JUN-2001; 2001WO-US020171.
>
> 26-JUN-2000; 2000US-0214047P.
> (SUCE-) SUGEN INC.
>
> Plowman G, Whyte D, Sudarsanam S, Manning G, Caenepeel S;
> Charyczak G;
> WPI; 2002-139913/18.
> N-PSDB; ABK31782.
>
> Nucleic acids encoding novel human proteases, useful for useful for
> treating diseases and disorders such as cancers, immune-related diseases
> and disorders, cardiovascular disease (e.g. restenosis) and inflammatory
> disorders.
>
> Claim 6; Fig 2N; 313pp; English.
>
> The present invention relates to the isolation of novel human proteases,
> and the nucleic acids encoding them. The sequences of the invention are
> useful for treating diseases and disorders such as cancers (e.g. breast,
> colon, lung), immune-related diseases and disorders (e.g. inflammatory
> diseases and asthma), cardiovascular diseases (e.g. restenosis and
> coronary thrombosis), brain or neuronal-associated diseases, metabolic
> disorders (e.g. diabetes, obesity), inflammatory disorders (e.g.
> rheumatoid arthritis and psoriasis), central or peripheral nervous system
> diseases, migraines, pain, sexual dysfunction, mood disorders, attention
> disorders, cognitive disorders, hypotension, hypertension, psychotic
> disorders, neurological disorders (e.g. Alzheimer's disease, Parkinson's
> disease) and dyskinesias. The nucleic acids and polypeptides are also
> useful for treating viral infections caused by human immunodeficiency
> virus (HIV), and non-viral infections such as ocular disease (e.g.
> glaucoma) and macular degeneration. AAU82702-AAU82760 represent the novel
> human proteases of the invention
>
> Sequence 253 AA;
>
> Query Match 100.0%; Score 40; DB 5; Length 253;
> Best Local Similarity 100.0%; Pred. No. 18;
> Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
>
> Y 1 LPLQLQLLL 9
> |||||
> b 6 LPLQLQLLL 14
>
> RESULT 10
> BU07440

```

```

ID XX ABU07440 standard; protein; 253 AA.
AC XX ABU07440;
XX XX
DT XX 28-JAN-2003 (first entry)
XX XX
DE XX Protein differentially regulated in prostate cancer #43.
XX XX
KW XX Prostate cancer; gene expression; differential regulation;
KW molecular marker; drug target; cancer detection; cancer diagnosis;
KW cancer staging; cancer grading; cancer assessing; cancer monitoring.
OS XX Homo sapiens.
XX XX WO20020281639-A2.
PN XX
PD XX 17-OCT-2002.
XX XX
XX XX 08-APR-2002; 2002WO-US010824.
XX XX
PR XX 06-APR-2001; 2001US-0281731P.
PR XX 06-APR-2001; 2001US-0281732P.
XX XX
PA XX (ORIG-) ORIGENE TECHNOLOGIES INC.
XX XX
PI XX Sun 2, Jay G;
XX XX
DR XX WPI; 2003-058520/05.
XX XX N-PSDB; ABX10343.
XX XX
PT XX Novel genes which are differentially regulated in prostate cancer, useful
PT for diagnosing prostate cancer in prostate tissue sample and assessing
PT therapeutic or preventive intervention in prostate cancer patients.
XX XX
PS XX Claim 1; Page 293-294; 416pp; English.
XX XX
XX XX The invention describes genes (I) which are differentially regulated in
XX prostate cancer. (I) Is useful for diagnosing a prostate cancer in a
XX sample comprising prostate tissue, which involves determining the number
XX of target genes which are differentially-regulated in the sample, where
XX the number is indicative of the probability that the sample comprises
XX prostate cancer. (I) Is useful for assessing a therapeutic or preventive
XX intervention in a subject having a prostate cancer, which involves
XX determining the expression levels in a sample comprising prostate tissue
XX of target genes which are differentially-regulated in prostate cancer.
XX Preferably, the expression levels of at least 10 genes are determined.
XX (I) is also useful for identifying agents that modulate a biological
XX activity of a polypeptide differentially-regulated in prostate cancer
XX cells, which involves contacting a polypeptide differentially-regulated
XX in prostate cancer cells with a test agent under conditions effective for
XX the test agent to modulate a biological activity of the polypeptide, and
XX determining whether the test agent modulates the biological activity. (I)
XX is useful as molecular markers, as drug targets, and for detecting,
XX diagnosing, staging, grading, assessing, monitoring, prognosticating,
XX preventing or treating, determining predisposition to diseases and
XX conditions especially relating to prostate cancer. (I) and its expression
XX products are used in the diagnostic test to assay for presence of cancer
XX e.g., in tissue sections, in biopsy sample, in total RNA, in lymph, in
XX blood etc. (I) is useful for assessing cancer e.g., to determine the type
XX of cancer, its stage of development, the nature of genetic defect, etc.
XX The polypeptide encoded by (I) can be used as target for therapy or drug
XX discovery. (I) can also be used for expressing the polypeptide, and thus
XX for searching specific binding partners of the polypeptide. (I) is useful
XX in therapeutic applications to treat prostate cancer. The identification
XX of specific genes, and groups of genes, expressed in pathways
XX physiologically relevant to prostate cancer permits the definition of
XX functional and disease pathways and the delineation of targets in these
XX pathways which are useful in diagnostic, therapeutic, and clinical
XX applications. This is the amino acid sequence of a protein differentially
XX regulated in prostate cancer
XX
XX Sequence 253 AA;

```

Query Match 100.0%; Score 40; DB 6; Length 253;  
Best Local Similarity 100.0%; Pred. No. 18;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 LLPLQILL 9  
b 6 LLPLQILL 14

RESULT 11  
BU07471  
D ABU07471 standard; protein; 253 AA.  
X C ABU07471;  
X 28-JAN-2003 (first entry)  
X Protein differentially regulated in prostate cancer #74.  
X  
X Prostate cancer; gene expression; differential regulation;  
W molecular marker; drug target; cancer detection; cancer diagnosis;  
W cancer staging; cancer grading; cancer assessing; cancer monitoring.  
W  
X Homo sapiens.  
X  
X WO200281638-A2.  
X  
X 17-OCT-2002.  
X  
X 08-APR-2002; 2002WO-US010824.  
X  
X 06-APR-2001; 2001US-0281731P.  
X 06-APR-2001; 2001US-0281732P.  
X (ORIG-) ORIGENE TECHNOLOGIES INC.  
X  
X Sun Z, Jay G;  
X  
X WPI; 2003-058520/05.  
X N-PSDB; ABX10375.  
X  
X Novel genes which are differentially regulated in prostate cancer, useful  
T for diagnosing prostate cancer in prostate tissue sample and assessing  
T therapeutic or preventive intervention in prostate cancer patients.  
T  
X Claim 1; Page 351; 416pp; English.  
X  
X The invention describes genes (I) which are differentially regulated in  
X prostate cancer. (I) is useful for diagnosing a prostate cancer in a  
X sample comprising prostate tissue, which involves determining the number  
X of target genes which are differentially-regulated in the sample, where  
X the number is indicative of the probability that the sample comprises  
X prostate cancer. (I) is useful for assessing a therapeutic or preventive  
X intervention in a subject having a prostate cancer, which involves  
X determining the expression levels in a sample comprising prostate tissue  
X of target genes which are differentially-regulated in prostate cancer.  
X Preferably, the expression levels of at least 10 genes are determined.  
X (I) is also useful for identifying agents that modulate a biological  
X activity of a polypeptide differentially-regulated in prostate cancer  
X cells, which involves contacting a polypeptide differentially-regulated  
X in prostate cancer cells with a test agent under conditions effective for  
X the test agent to modulate a biological activity of the polypeptide, and  
X determining whether the test agent modulates the biological activity. (I)  
X is useful as molecular markers, as drug targets, and for detecting,  
X diagnosing, staging, grading, assessing, monitoring, prognosticating,  
X preventing or treating, determining predisposition to diseases and  
X conditions especially relating to prostate cancer. (I) and its expression  
X products are used in the diagnostic test to assay for presence of cancer  
X e.g., in tissue sections, in biopsy sample, in total RNA, in lymph, in  
X blood etc. (I) is useful for assessing cancer e.g., to determine the type  
X of cancer, its stage of development, the nature of genetic defect, etc.  
X The polypeptide encoded by (I) can be used as target for therapy or drug  
X discovery. (I) can also be used for expressing the polypeptide and thus

CC for searching specific binding partners of the polypeptide. (I) is useful  
CC in therapeutic applications to treat prostate cancer. The identification  
CC of specific genes, and groups of genes, expressed in pathways  
CC physiologically relevant to prostate cancer permits the definition of  
CC functional and disease pathways and the delineation of targets in these  
CC pathways which are useful in diagnostic, therapeutic, and clinical  
CC applications. This is the amino acid sequence of a protein differentially  
CC regulated in prostate cancer  
X  
X SQ Sequence 253 AA;

Query Match 100.0%; Score 40; DB 6; Length 253;  
Best Local Similarity 100.0%; Pred. No. 18;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LLPLQILL 9  
Db 6 LLPLQILL 14

RESULT 12  
ABR58471  
ID ABR58471 standard; protein; 253 AA.

XX ABR58471;  
XX  
XX 07-JUL-2003 (first entry)  
XX Human stratum corneum chymotryptic enzyme - ovarian cancer clone O1676P.  
XX Human; cytostatic; gene therapy; vaccine; cancer; ovarian cancer.  
XX Homo sapiens.  
XX WO2003029468-A1.  
XX 10-APR-2003.  
XX 02-OCT-2002; 2002WO-US031467.  
XX  
XX 02-OCT-2001; 2001US-0327135P.  
XX 30-MAY-2002; 2002US-0384531P.  
XX (CORI-) CORIXA CORP.  
XX  
XX Algate PA, Mannion J;  
XX WPI; 2003-372001/35.  
XX  
XX New polynucleotide and polypeptide useful for diagnosing and/or treating  
PT cancer, particularly ovarian cancer, and as a vaccine.  
XX  
XX Claim 2; Page 157-158; 169pp; English.  
XX  
XX The invention relates to a novel isolated polynucleotide. The  
CC polynucleotides of the invention have cytostatic activity, and may have a  
CC use in gene therapy, and in a vaccine. The composition and methods are  
CC useful in diagnosing and/or treating cancer, particularly ovarian cancer.  
CC The composition may also be used as a vaccine to prevent cancer. The  
CC present sequence is used in the exemplification of the invention  
X  
X SQ Sequence 253 AA;

Query Match 100.0%; Score 40; DB 6; Length 253;  
Best Local Similarity 100.0%; Pred. No. 18;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LLPLQILL 9  
Db 6 LLPLQILL 14

RESULT 13

DB80484  
 D ADB80484 standard; protein; 253 AA.  
 X  
 C ADB80484;  
 C  
 K 04-DEC-2003 (first entry)  
 K  
 K Ovarian cancer-associated protein #24.  
 E  
 K cytostatic; gene therapy; vaccine; ovarian cancer; diagnosis;  
 K post-operative chemotherapy; radiation therapy; tumour prognosis;  
 W pre-cancerous lesion detection.  
 W  
 K Homo sapiens.  
 S  
 K WO2002102235-A2.  
 V  
 K 27-DEC-2002.  
 C  
 K 18-JUN-2002; 2002WO-US019297.  
 K  
 K 18-JUN-2001; 2001US-0299234P.  
 K  
 R 27-AUG-2001; 2001US-0315287P.  
 R  
 R 05-SEP-2001; 2001US-0317544P.  
 R  
 R 13-NOV-2001; 2001US-0350666P.  
 R  
 R 12-APR-2002; 2002US-0372248P.  
 X  
 A (EOSB-) EOS BIOTECHNOLOGY INC.  
 X  
 X Mack DH, Gish KC;  
 I  
 I WPI; 2003-167431/16.  
 X  
 R N-PSDB; ADB80483.  
 X  
 T Detecting an ovarian cancer-associated transcript in a cell from a  
 T patient, comprises contacting a biological sample from the patient with a  
 T polynucleotide that hybridizes to an ovarian cancer gene.  
 X  
 S Claim 13; Page 291; 332pp; English.  
 X  
 X The invention relates to a method of detecting an ovarian cancer-  
 C associated transcript in a cell from a patient, by contacting a  
 C biological sample from the patient with a polynucleotide that selectively  
 C hybridizes to a sequence at least 80% identical to any of one of 80  
 C nucleic acid sequences given in the specification. The method is useful  
 C in diagnosing ovarian cancer and in identifying and using agents and/or  
 C targets that inhibit ovarian cancer. The nucleic acid molecule,  
 C polypeptide and the antibody may also be used in detecting ovarian  
 C cancers, monitoring and early detection of relapse following treatment,  
 C monitoring response to therapy, selecting patients for post-operative  
 C chemotherapy or radiation therapy, in selecting mode of therapy,  
 C determining tumour prognosis, early detection of pre-cancerous lesions,  
 C and as vaccines. This sequence corresponds to one of the proteins used  
 C for the detection method of the invention.  
 X  
 C Sequence 253 AA;  
 Q

XX Human HSCE.  
 DE  
 XX  
 XX Human; KLK-L1; KLK-L2; KLK-L3; KLK-L4; KLK-L5; KLK-L6; HSCE;  
 KW human stratum corneum chymotryptic enzyme; kallikrein-like protein;  
 KW serine protease; cytostatic; cancer; prostrate cancer.  
 XX  
 OS Homo sapiens.  
 OS  
 PN WO200053776-A2.  
 PN  
 PD 14-SEP-2000.  
 XX  
 PF 09-MAR-2000; 2000WO-CA000258.  
 XX  
 XX 11-MAR-1999; 99US-0124260P.  
 PR  
 PR 01-APR-1999; 99US-0127386P.  
 PR  
 PR 21-JUL-1999; 99US-0144919P.  
 XX  
 XX (MOUN) MOUNT SINAI HOSPITAL.  
 PA  
 XX  
 XX Yousef GM, Diamandis EP;  
 PI  
 XX WPI; 2000-587440/55.  
 DR  
 XX  
 XX New kallikrein-like (KLK-L) proteins for diagnosing and treating KLK-L  
 PT protein mediated disorders, especially cancer.  
 PT  
 XX Example 4; Fig 17; 184pp; English.  
 PS  
 XX The present sequence is human stratum corneum chymotryptic enzyme  
 CC (HSCE), a member of the kallikrein multi-gene family. Kallikreins and  
 CC kallikrein-like proteins are a subgroup of the serine protease enzyme  
 CC family. They catalyse the selective cleavage of specific polypeptide  
 CC precursors to release peptides with potent biological activity. Nucleic  
 CC acids encoding kallikrein-like proteins KLK-L1, KLK-L2, KLK-L3, KLK-L4,  
 CC KLK-L5 and KLK-L6 have been isolated. The proteins are useful in the  
 CC treatment, monitoring and diagnosis of cancers, especially prostate  
 CC cancer. They can also be used to identify a substance that can associate  
 CC with or mediate the biological activity of the proteins. Antibodies can  
 CC be used to treat conditions mediated by the kallikrein-like proteins  
 XX  
 SQ Sequence 257 AA;  
 Q

Query Match 100.0%; Score 40; DB 3; Length 257;  
 Best Local Similarity 100.0%; Pred. No. 18;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLPLQLLLL 9  
 Db 6 LLPLQLLLL 14  
 XX

RESULT 15  
 AAB63580  
 ID AAB63580 standard; protein; 142 AA.  
 XX  
 AC AAB63580;  
 XX  
 XX 26-MAR-2001 (first entry)  
 XX  
 DE Human gastric cancer associated antigen protein sequence SEQ ID NO:942.  
 XX Human; breast cancer; gastric cancer; prostate cancer; diagnosis;  
 KW cancer associated antigen; cytostatic; cancer vaccine.  
 XX  
 OS Homo sapiens.  
 OS  
 XX WO2000073801-A2.  
 PN  
 XX 07-DEC-2000.  
 PD  
 XX 26-MAY-2000; 2000WO-US014749.  
 PF

X 28-MAY-1999; 99US-0136526P.  
R 10-SEP-1999; 99US-0153454P.  
X A (LUDW-) LUDWIG INST CANCER RES.  
X I Obata Y;  
X R WPI; 2001-025274/03.  
X T Nucleic acids encoding breast, gastric and prostate cancer associated  
T antigen precursors, useful for diagnosing and treating a condition  
T characterized by expression of an abnormal amount of a protein, e.g.  
T cancer.  
S Example 1; Page 625; 799pp; English.  
X AAF22422 to AAF22626, AAF22627 to AAF22773 and AAF22774 to AAF23014  
C represent nucleotide sequences encoding human breast, gastric and  
C prostate cancer associated antigen precursors (CAAP) respectively.  
C AAB63232 to AAB63467, AAB63468 to AAB63721 and AAB63722 to AAB63970  
C represent human breast, gastric and prostate CAAP protein sequence  
C respectively. CAAPs have cytostatic activity and can be used in the  
C production of cancer vaccines. The human CAAP proteins, peptides, nucleic  
C acids or anti-CAAP antibodies are useful for diagnosing and treating a  
C condition characterised by expression of an abnormal amount of a protein,  
C e.g. cancer  
X Q Sequence 142 AA;  
Query Match 95.0%; Score 38; DB 4; Length 142;  
Best Local Similarity 88.9%; Pred. No. 23;  
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
Y 1 LLPLQLLL 9  
b 48 LLPLQLLL 56  
earch completed: March 1, 2004, 17:28:52  
ob time : 46.5556 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

4 protein - protein search, using sw model

in on: March 1, 2004, 17:21:46 ; Search time 11.8889 Seconds  
(without alignments)  
39.081 Million cell updates/sec

itle: US-09-905-083-35

arfect score: 40

equence: 1 LLPLQLILL 9

oring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

sarched: 389414 seqs, 51625971 residues

otal number of hits satisfying chosen parameters: 389414

inimum DB seq length: 0

aximum DB seq length: 2000000000

ost-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

atabase :

Issued Patents AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A-COMB.pap:\*  
2: /cgn2\_6/ptodata/2/iaa/5B-COMB.pap:\*  
3: /cgn2\_6/ptodata/2/iaa/6A-COMB.pap:\*  
4: /cgn2\_6/ptodata/2/iaa/6B-COMB.pap:\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS-COMB.pap:\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

result No.	Score	Query Match %	Length	DB ID	Description
1	40	100.0	9	3	US-09-502-600-35
2	40	100.0	9	4	US-09-918-243-35
3	40	100.0	253	2	US-08-557-146-2
4	40	100.0	253	2	US-08-824-874-3
5	40	100.0	253	2	US-09-154-344-2
6	40	100.0	253	3	US-08-930-188-2
7	40	100.0	253	3	US-09-210-084-3
8	40	100.0	253	3	US-09-764-762-3
9	40	100.0	253	5	PCT-US96-04294-2
10	36	90.0	9	3	US-09-502-600-33
11	36	90.0	73	4	US-09-918-243-33
12	34	85.0	73	4	US-09-205-258-1083
13	34	85.0	302	4	US-09-393-634-37
14	34	85.0	424	4	US-09-205-258-1079
15	34	85.0	569	4	US-09-252-991A-22870
16	34	85.0	812	4	US-09-489-039A-12075
17	33	82.5	190	1	US-08-339-152A-19
18	33	82.5	190	2	US-08-007-999B-6
19	33	82.5	190	2	US-08-689-276A-6
20	33	82.5	232	2	US-08-956-047-36
21	33	82.5	634	1	US-08-339-152A-17
22	33	82.5	653	1	US-08-339-152A-16
23	33	82.5	653	2	US-08-007-999B-3
24	33	82.5	653	2	US-08-689-276A-3
25	33	82.5	2787	3	US-09-245-041-15
26	32	80.0	9	3	US-09-502-600-36
27	32	80.0	9	4	US-09-918-243-36

Sequence 28, Appl  
Sequence 28, Appl  
Sequence 42, Appl  
Sequence 38, Appl  
Sequence 36, Appl  
Sequence 9507, Ap  
Sequence 2, Appl  
Sequence 2, Appl  
Sequence 2, Appl  
Sequence 1, Appl  
Sequence 2, Appl  
Sequence 4, Appl  
Sequence 4, Appl  
Sequence 4, Appl  
Sequence 5063, Ap  
Sequence 4, Appl  
Sequence 4, Appl

28 32 80.0 12 3 US-09-502-600-28  
29 32 80.0 12 4 US-09-918-243-28  
30 32 80.0 264 3 US-08-713-556F-42  
31 32 80.0 269 3 US-08-713-556F-38  
32 32 80.0 313 3 US-08-713-556F-36  
33 32 80.0 395 4 US-09-489-039A-9507  
34 32 80.0 402 2 US-08-477-254A-2  
35 32 80.0 402 2 US-08-472-576B-2  
36 32 80.0 402 2 US-08-428-734B-2  
37 32 80.0 402 3 US-09-063-237-1  
38 32 80.0 402 3 US-08-713-556F-2  
39 32 80.0 412 2 US-08-477-254A-4  
40 32 80.0 412 2 US-08-472-576B-4  
41 32 80.0 412 2 US-08-428-734B-4  
42 32 80.0 412 3 US-08-713-556F-4  
43 32 80.0 484 4 US-09-134-001C-5063  
44 32 80.0 492 1 US-07-794-393-4  
45 32 80.0 492 1 US-08-001-711-4

## ALIGNMENTS

### RESULT 1

US-09-502-600-35

; Sequence 35, Application US/09502600A

; Patent No. 6294344

; GENERAL INFORMATION:

; APPLICANT: O'Brien, Timothy J.

; TITLE OF INVENTION: Compositions and Methods for the Early Diagnosis of

; FILE REFERENCE: D6223CIP-C

; CURRENT FILING DATE: 2000-02-11

; CURRENT APPLICATION NUMBER: US/09/502,600A

; PRIOR APPLICATION NUMBER: 09/039,211

; PRIOR FILING DATE: 03-14-1998

; NUMBER OF SEQ ID NOS: 136

; SEQ ID NO 35

; LENGTH: 9

; TYPE: PRT

; ORGANISM: Homo sapiens

; FEATURE:

; OTHER INFORMATION: Residues 6-14 of the SCCE protein

US-09-502-600-35

Query Match 100.0%; Score 40; DB 3; Length 9;

Best Local Similarity 100.0%; Pred. No. 3e+05; 0; Mismatches 0; Gaps 0;

Matches 9; Conservative 0; Indels 0; Gaps 0;

Qy 1 LLPLQLILL 9

Db 1 LLPLQLILL 9

### RESULT 2

US-09-918-243-35

; Sequence 35, Application US/09918243

; Patent No. 6627403

; GENERAL INFORMATION:

; APPLICANT: O'Brien, Timothy J.

; APPLICANT: Cannon, Martin J.

; APPLICANT: Santin, Alessandro

; TITLE OF INVENTION: Methods for the early diagnosis of ovarian cancer

; FILE REFERENCE: D6223CIP/C/D/CIP

; CURRENT APPLICATION NUMBER: US/09/918,243

; CURRENT FILING DATE: 2001-07-30

; PRIOR APPLICATION NUMBER: US

; PRIOR FILING DATE: 2001-07-13

; NUMBER OF SEQ ID NOS: 136

; SEQ ID NO 35

; LENGTH: 9

; TYPE: PRT

; ORGANISM: Homo sapiens

## FEATURE:

NAME/KEY: CHAIN

OTHER INFORMATION: Residues 6-14 of the SCCE protein

3-09-918-243-35

Query Match 100.0%; Score 40; DB 4; Length 9;

Best Local Similarity 100.0%; Pred. No. 3e-05;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 LLPLQILL 9

D 1 LLPLQILL 9

## RESULT 3

S-08-557-146-2

Sequence 2, Application US/08557146

Patent No. 5834290

GENERAL INFORMATION:

APPLICANT: Egelrud, Torbjorn

APPLICANT: Hanson, Lennart

TITLE OF INVENTION: Recombinant Stratum Corneum Chymotryptic

TITLE OF INVENTION: Enzyme (SCCE)

NUMBER OF SEQUENCES: 17

CORRESPONDENCE ADDRESS:

ADDRESSEE: White &amp; Case, Patent Department

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: U.S.A.

ZIP: 10036-2787

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/557,146

FILING DATE: 14-DEC-1995

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: Steiner, Richard J.

REGISTRATION NUMBER: 35,372

REFERENCE/DOCKET NUMBER: 1103326-181

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 819-8783

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 253 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

S-08-557-146-2

Query Match

Best Local Similarity 100.0%; Score 40; DB 2; Length 253;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 LLPLQILL 9

D 6 LLPLQILL 14

## RESULT 4

S-08-824-874-3

Sequence 3, Application US/08824874

Patent No. 5962300

GENERAL INFORMATION:

APPLICANT: Hillman, Jennifer L.

APPLICANT: Lal, Preeti

TITLE OF INVENTION: NOVEL KALLIKREIN

NUMBER OF SEQUENCES: 5

## CORRESPONDENCE ADDRESS:

ADDRESSEE: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Drive

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/824,874

FILING DATE: Filed Herewith

CLASSIFICATION: 514

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.

REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PF-0252 US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-855-0555

TELEFAX: 415-845-4166

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 253 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

LIBRARY: GenBank

CLONE: 532504

US-08-824-874-3

Query Match

Best Local Similarity 100.0%; Score 40; DB 2; Length 253;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLPLQILL 9

Db 6 LLPLQILL 14

## RESULT 5

US-09-154-344-2

Sequence 2, Application US/09154344

Patent No. 5981256

GENERAL INFORMATION:

APPLICANT: Egelrud, Torbjorn

APPLICANT: Hanson, Lennart

TITLE OF INVENTION: Recombinant Stratum Corneum Chymotryptic

TITLE OF INVENTION: Enzyme (SCCE)

NUMBER OF SEQUENCES: 17

CORRESPONDENCE ADDRESS:

ADDRESSEE: White &amp; Case, Patent Department

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: U.S.A.

ZIP: 10036-2787

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/154,344

FILING DATE: 16-SEP-1998

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/557,146  
FILING DATE: 14-DEC-1995  
CLASSIFICATION:  
NAME: Steiner, Richard J.  
REGISTRATION NUMBER: 35,372  
REFERENCE/DOCKET NUMBER: 1103326-181  
TELEPHONE: (212) 819-8783  
TELEFAX: (212) 354-8113  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 253 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
S-09-154-344-2

Query Match 100.0%; Score 40; DB 2; Length 253;  
Best Local Similarity 100.0%; Pred. No. 3;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 LLPLQILL 9  
b 6 LLPLQILL 14

RESULT 6  
IS-08-930-188-2  
Sequence 2, Application US/08930188  
Patent No. 6093397  
GENERAL INFORMATION:  
APPLICANT: Dixon, Eric P.  
APPLICANT: Johnstone, Edward M.  
APPLICANT: Little, Sheila P.  
TITLE OF INVENTION: AMYLOID PRECURSOR PROTEIN PROTEASE AND  
RELATED NUCLEIC ACIDS  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Eli Lilly and Company  
STREET: Lilly Corporate Center  
CITY: Indianapolis  
STATE: Indiana  
COUNTRY: United States of America  
ZIP: 46285  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/930,188  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/416,257  
FILING DATE: 04-APR-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Blalock, Donna K.  
REGISTRATION NUMBER: 38,082  
REFERENCE/DOCKET NUMBER: X9239  
TELEPHONE: 317-277-1090  
TELEFAX: 317-276-3861  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 253 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
IS-08-930-188-2

Query Match 100.0%; Score 40; DB 3; Length 253;  
Best Local Similarity 100.0%; Pred. No. 3;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LLPLQILL 9  
Db 6 LLPLQILL 14

RESULT 7  
US-09-210-084-3  
Sequence 3, Application US/09210084  
Patent No. 6197511  
GENERAL INFORMATION:  
APPLICANT: Hillman, Jennifer L.  
APPLICANT: Lal, Preeti  
TITLE OF INVENTION: NOVEL KALLIKREIN  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/210,084  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/824,874  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0252 US  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 253 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 532504  
US-09-210-084-3

Query Match 100.0%; Score 40; DB 3; Length 253;  
Best Local Similarity 100.0%; Pred. No. 3;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LLPLQILL 9  
Db 6 LLPLQILL 14

RESULT 8  
US-09-764-762-3  
Sequence 3, Application US/09764762  
Patent No. 6472195  
GENERAL INFORMATION:  
APPLICANT: Hillman, Jennifer L.  
APPLICANT: Lal, Preeti  
TITLE OF INVENTION: NOVEL KALLIKREIN  
NUMBER OF SEQUENCES: 5

CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/764,762  
FILING DATE: 16-Jan-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/210,084  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0252 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 253 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 532504  
SEQUENCE DESCRIPTION: SEQ ID NO: 3:  
S-09-764-762-3  
Query Match 100.0%; Score 40; DB 4; Length 253;  
Best Local Similarity 100.0%; Pred. No. 3;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 LLPLQLILL 9  
| | | | |  
b 6 LLPLQLILL 14

RESULT 9  
CT-US96-04294-2  
Sequence 2, Application PC/TUS9604294  
GENERAL INFORMATION:  
APPLICANT: Dixon, Eric P.  
APPLICANT: Johnstone, Edward M.  
APPLICANT: Little, Sheila P.  
TITLE OF INVENTION: AMYLOID PRECURSOR PROTEIN PROTEASE AND  
TITLE OF INVENTION: RELATED NUCLEIC ACIDS  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Eli Lilly and Company  
STREET: Lilly Corporate Center  
CITY: Indianapolis  
STATE: Indiana  
COUNTRY: United States of America  
ZIP: 46285  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/04294  
FILING DATE:  
CLASSIFICATION:

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/416,257  
FILING DATE: 04-APR-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Blalock, Donna K.  
REGISTRATION NUMBER: 38,082  
REFERENCE/DOCKET NUMBER: X9239  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 317-277-1090  
TELEFAX: 317-276-3861  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 253 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US96-04294-2  
Query Match 100.0%; Score 40; DB 5; Length 253;  
Best Local Similarity 100.0%; Pred. No. 3;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLPLQLILL 9  
| | | | |  
Db 6 LLPLQLILL 14

RESULT 10  
US-09-502-600-33  
Sequence 33, Application US/09502600A  
Patent No. 6294344  
GENERAL INFORMATION:  
APPLICANT: O'Brien, Timothy J.  
TITLE OF INVENTION: Compositions and Methods for the Early Diagnosis of  
TITLE OF INVENTION: Ovarian Cancer  
FILE REFERENCE: D6223CIP-C  
CURRENT FILING DATE: 2000-02-11  
CURRENT APPLICATION NUMBER: US/09/502,600A  
PRIOR APPLICATION NUMBER: 09/039,211  
PRIOR FILING DATE: 03-14-1998  
NUMBER OF SEQ ID NOS: 136  
SEQ ID NO 33  
LENGTH: 9  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: Residues 5-13 of the SCCE protein  
US-09-502-600-33

Query Match 90.0%; Score 36; DB 3; Length 9;  
Best Local Similarity 100.0%; Pred. No. 3e+05;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLPLQLILL 8  
| | | | |  
Db 2 LLPLQLILL 9

RESULT 11  
US-09-918-243-33  
Sequence 33, Application US/09918243  
Patent No. 6627403  
GENERAL INFORMATION:  
APPLICANT: O'Brien, Timothy J.  
APPLICANT: Cannon, Martin J.  
APPLICANT: Santin, Alessandro  
TITLE OF INVENTION: Methods for the early diagnosis of ovarian cancer  
FILE REFERENCE: D6223CIP/C/D/CIP  
CURRENT APPLICATION NUMBER: US/09/918,243  
CURRENT FILING DATE: 2001-07-30  
PRIOR APPLICATION NUMBER: US  
PRIOR FILING DATE: 2001-07-13

NUMBER OF SEQ ID NOS: 136  
SEQ ID NO 33  
LENGTH: 9  
TYPE: PRT  
ORGANISM: Homo sapiens  
NAME/KEY: CHAIN  
FEATURE:  
OTHER INFORMATION: Residues 5-13 of the SCCE protein  
S-09-918-243-33

Query Match 90.0%; Score 36; DB 4; Length 9;  
Best Local Similarity 100.0%; Pred. No. 3e+05; 0; Indels 0;  
Matches 8; Conservative 0; Mismatches 0; Gaps 0;

Y 1 LPLQLILL 8  
|||  
b 2 LPLQLILL 9

RESULT 12  
S-09-205-258-1083  
Sequence 1083, Application US/09205258  
Patent No. 6525174  
GENERAL INFORMATION:  
APPLICANT: Young et al.  
TITLE OF INVENTION: 207 Human Secreted Proteins  
FILE REFERENCE: PZ007P1  
CURRENT APPLICATION NUMBER: US/09/205,258  
CURRENT FILING DATE: 1998-12-04  
EARLIER APPLICATION NUMBER: PCT/US98/11422  
EARLIER FILING DATE: 1998-06-04  
EARLIER APPLICATION NUMBER: 60/048,895  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/049,375  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,881  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,880  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,896  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/049,020  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,876  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,895  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,884  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,894  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,971  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,964  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,882  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,899  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,893  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,900  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,901  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,892  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,915  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/049,019  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,970  
EARLIER FILING DATE: 1997-06-06

EARLIER APPLICATION NUMBER: 60/048,972  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,916  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/049,373  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,875  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/049,374  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,917  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,949  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,974  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,883  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,897  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,898  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,962  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,963  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,877  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,878  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/070,923  
EARLIER FILING DATE: 1997-12-18  
EARLIER APPLICATION NUMBER: 60/092,921  
EARLIER FILING DATE: 1998-07-15  
EARLIER APPLICATION NUMBER: 60/094,657  
NUMBER OF SEQ ID NOS: 1227  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 1083  
LENGTH: 73  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (8)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-205-258-1083

Query Match 85.0%; Score 34; DB 4; Length 73;  
Best Local Similarity 87.5%; Pred. No. 11;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 LPLQLILL 9  
|||  
Db 25 LPLQLILL 32

RESULT 13  
US-09-393-634-37  
Sequence 37, Application US/09393634  
Patent No. 6558910  
GENERAL INFORMATION:  
APPLICANT: Zuker, Charles S.  
APPLICANT: Adler, Jon Elliot  
APPLICANT: Ryba, Nick  
APPLICANT: Mueller, Ken  
APPLICANT: Hoon, Mark  
APPLICANT: The Regents of the University of California  
APPLICANT: The Government of the United States of America  
APPLICANT: as represented by the Secretary of the  
APPLICANT: Department of Health and Human Services  
TITLE OF INVENTION: SF, a No. 6558910el Family of Taste Receptors  
FILE REFERENCE: 02307E-098000US

CURRENT APPLICATION NUMBER: US/09/393,634  
CURRENT FILING DATE: 1998-09-10  
NUMBER OF SEQ ID NOS: 92  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 37  
LENGTH: 302  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: human GR02  
NAME/KEY: MOD RES  
LOCATION: (143)  
OTHER INFORMATION: Xaa = any amino acid  
S-09-393-634-37

Query Match 85.0%; Score 34; DB 4; Length 302;  
Best Local Similarity 66.7%; Pred. No. 49;  
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

y 1 LPLQILL 9  
|:|:|:|:  
b 42 LMPIQILL 50

RESULT 14  
S-09-205-258-1079  
Sequence 1079, Application US/09205258  
Patent No. 6525174  
GENERAL INFORMATION:  
APPLICANT: Young et al.  
TITLE OF INVENTION: 207 Human Secreted Proteins  
FILE REFERENCE: P2007P1  
CURRENT APPLICATION NUMBER: US/09/205,258  
CURRENT FILING DATE: 1998-12-04  
EARLIER APPLICATION NUMBER: PCT/US98/11422  
EARLIER FILING DATE: 1998-06-04  
EARLIER APPLICATION NUMBER: 60/048,885  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/049,375  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,881  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,880  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,896  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/049,020  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,876  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,895  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,884  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,894  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,971  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,964  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,882  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,899  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,893  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,900  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,901  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,892  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,915

; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/049,019  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,970  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,972  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,916  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/049,373  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,875  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/049,374  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,917  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,949  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,974  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,883  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,897  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,898  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,962  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,963  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,877  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,878  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/070,923  
; EARLIER FILING DATE: 1997-12-18  
; EARLIER APPLICATION NUMBER: 60/092,921  
; EARLIER FILING DATE: 1998-07-15  
; EARLIER APPLICATION NUMBER: 60/094,657  
; EARLIER FILING DATE: 1998-07-30  
; NUMBER OF SEQ ID NOS: 1227  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1079  
; LENGTH: 424  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (152)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (314)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (359)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; US-09-205-258-1079

Query Match 85.0%; Score 34; DB 4; Length 424;  
Best Local Similarity 87.5%; Pred. No. 70;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LPLQILL 9  
|:|:|:|:  
Db 376 LPLQILL 383

RESULT 15  
US-09-252-991A-22870  
; Sequence 22870, Application US/09252991A

Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 22870  
LENGTH: 569  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
;-09-252-991A-22870

Query Match 85.0%; Score 34; DB 4; Length 569;  
Best Local Similarity 77.8%; Pred. No. 95;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
/ 1 LLPLQILL 9  
|||:|  
51 LLPLSVLL 59

Search completed: March 1, 2004, 17:38:24  
Search time : 11.8889 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

M protein - protein search, using sw model

un on: March 1, 2004, 17:35:01 ; Search time 24.1111 Seconds  
(without alignments)  
78.818 Million cell updates/sec

title: US-09-905-083-35  
effect score: 40  
equene: 1 LLPQLILL 9

coring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

searched: 809742 seqs, 211153259 residues

total number of hits satisfying chosen parameters: 809742

inimum DB seq length: 0  
aximum DB seq length: 2000000000  
ost-processing: Minimum Match 0%  
Listing first 45 summaries

atabase : Published Applications AA.\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/1/pubpaa/ECT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/1/pubpaa/ECTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	ID	Description
1	40	100.0	9	9	US-09-918-243-35
2	40	100.0	9	9	US-09-905-083-35
3	40	100.0	253	9	US-09-888-615-98
4	40	100.0	253	9	US-09-764-762-3
5	40	100.0	253	14	US-10-264-283-90
6	40	100.0	253	15	US-10-295-027-498
7	40	100.0	253	15	US-10-173-999-48
8	36	90.0	9	9	US-09-918-243-33
9	36	90.0	9	9	US-09-905-083-33
10	36	90.0	868	15	US-10-369-493-10255
11	35	87.5	23	9	US-09-864-761-46097
12	35	87.5	480	9	US-09-895-913A-92
13	34	85.0	70	14	US-10-097-065-220
14	34	85.0	73	10	US-09-933-767-1083
15	34	85.0	73	14	US-10-023-282-1083

16	34	85.0	183	15	US-10-264-237-2472	Sequence 2472, Ap
17	34	85.0	201	10	US-09-956-622A-39	Sequence 39, Appl
18	34	85.0	302	9	US-09-393-634-37	Sequence 37, Appl
19	34	85.0	302	10	US-09-510-332-3	Sequence 37, Appl
20	34	85.0	302	14	US-10-383-982-37	Sequence 1079, Ap
21	34	85.0	424	10	US-09-933-767-1079	Sequence 1079, Ap
22	34	85.0	424	14	US-10-023-282-1081	Sequence 1081, A
23	34	85.0	559	15	US-10-156-761-10881	Sequence 10881, A
24	34	85.0	1245	15	US-10-369-493-20447	Sequence 20447, A
25	34	85.0	1993	13	US-10-098-979-2	Sequence 2, Appl
26	33	82.5	146	15	US-10-291-285-881	Sequence 881, App
27	33	82.5	181	15	US-10-291-285-409	Sequence 409, App
28	33	82.5	201	9	US-09-989-722-99	Sequence 99, Appl
29	33	82.5	201	9	US-09-989-723-99	Sequence 99, Appl
30	33	82.5	201	9	US-09-989-727-99	Sequence 99, Appl
31	33	82.5	201	9	US-09-989-727-99	Sequence 99, Appl
32	33	82.5	201	9	US-09-989-731-99	Sequence 99, Appl
33	33	82.5	201	9	US-09-989-732-99	Sequence 99, Appl
34	33	82.5	201	9	US-09-991-073-99	Sequence 99, Appl
35	33	82.5	201	9	US-09-990-442-99	Sequence 99, Appl
36	33	82.5	201	9	US-09-991-163-99	Sequence 99, Appl
37	33	82.5	201	9	US-09-993-604-99	Sequence 99, Appl
38	33	82.5	201	9	US-09-990-456-99	Sequence 99, Appl
39	33	82.5	201	9	US-09-989-721-99	Sequence 99, Appl
40	33	82.5	201	9	US-09-992-598-99	Sequence 99, Appl
41	33	82.5	201	9	US-09-984-245-136	Sequence 136, App
42	33	82.5	201	9	US-09-989-293A-99	Sequence 99, Appl
43	33	82.5	201	9	US-09-989-735-99	Sequence 99, Appl
44	33	82.5	201	9	US-09-990-444-99	Sequence 99, Appl
45	33	82.5	201	9	US-09-991-181-99	Sequence 99, Appl

ALIGNMENTS

RESULT 1

US-09-918-243-35  
; Sequence 35, Application US/09918243  
; Parent No. US20020142317A1  
; GENERAL INFORMATION:  
; APPLICANT: O'Brien, Timothy J.  
; APPLICANT: Cannon, Martin J.  
; APPLICANT: Santin, Alessandro  
; TITLE OF INVENTION: Methods for the early diagnosis of ovarian cancer  
; FILE REFERENCE: D6223CIP/C/D/CIP  
; CURRENT APPLICATION NUMBER: US/09/918,243  
; CURRENT FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US  
; PRIOR FILING DATE: 2001-07-13  
; NUMBER OF SEQ ID NOS: 136  
; SEQ ID NO 35  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CHAIN  
; OTHER INFORMATION: Residues 6-14 of the SCCE protein  
US-09-918-243-35

Query Match 100.0%; Score 40; DB 9; Length 9;  
Best Local Similarity 100.0%; Pred No. 7.1e+05;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LLPQLILL 9  
Db 1 LLPQLILL 9

RESULT 2

US-09-905-083-35  
; Sequence 35, Application US/09905083  
; Patent No. US20020146708A1  
; GENERAL INFORMATION:



APPLICANT: O'Brien, Timothy J.  
TITLE OF INVENTION: Compositions and Methods for the Early Diagnosis of  
FILE OF INVENTION: Ovarian Cancer  
FILE REFERENCE: D6223CIP/C/DIV  
CURRENT APPLICATION NUMBER: US/09/905,083  
CURRENT FILING DATE: 2001-07-13  
PRIOR APPLICATION NUMBER: US 09/502,600  
PRIOR FILING DATE: 2000-02-11  
NUMBER OF SEQ ID NOS: 136  
SEQ ID NO 35  
LENGTH: 9  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CHAIN  
OTHER INFORMATION: Residues 6-14 of the SCCE protein

3-09-905-083-35

Query Match 100.0%; Score 40; DB 9; Length 9;  
Best Local Similarity 100.0%; Pred. No. 7.1e+05;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 LLPLQILL 9  
1 LLPLQILL 9

RESULT 3

3-09-888-615-98  
Sequence 98, Application US/09888615  
Patent No. US20020084856A1

GENERAL INFORMATION:

APPLICANT: FLOWMAN, GREGORY  
APPLICANT: WHYTE, DAVID  
APPLICANT: CAENEPEEL, SEAN  
APPLICANT: CHARYDCZAK, GLEN  
APPLICANT: MANNING, GERARD  
APPLICANT: SUDARSANAM, SUCHA

TITLE OF INVENTION: NOVEL PROTEASES

FILE REFERENCE: 038602/1214  
CURRENT APPLICATION NUMBER: US/09/888,615  
CURRENT FILING DATE: 2001-06-26  
PRIOR APPLICATION NUMBER: 60/214,047  
PRIOR FILING DATE: 2000-06-26  
NUMBER OF SEQ ID NOS: 150  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 98  
LENGTH: 253  
TYPE: PRT  
ORGANISM: Homo sapiens

3-09-888-615-98

Query Match 100.0%; Score 40; DB 9; Length 253;  
Best Local Similarity 100.0%; Pred. No. 18;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 LLPLQILL 9  
6 LLPLQILL 14

RESULT 4

3-09-764-762-3  
Sequence 3, Application US/09764762  
Patent No. US20020068341A1

GENERAL INFORMATION:

APPLICANT: Hillman, Jennifer L.

TITLE OF INVENTION: NOVEL KALLIKREIN

NUMBER OF SEQUENCES: 5

CORRESPONDENCE ADDRESS:

ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive

CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/764,762  
FILING DATE: 16-Jan-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/210,084  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0252 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 253 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 532504  
SEQUENCE DESCRIPTION: SEQ ID NO: 3:

US-09-764-762-3

Query Match 100.0%; Score 40; DB 9; Length 253;  
Best Local Similarity 100.0%; Pred. No. 18;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 LLPLQILL 9  
6 LLPLQILL 14

RESULT 5

US-10-264-283-90  
Sequence 90, Application US/10264283  
Publication No. US2003014494A1

GENERAL INFORMATION:

APPLICANT: Algate, Paul A.  
APPLICANT: Mannion, Jane  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
FILE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER  
FILE REFERENCE: 210121.590  
CURRENT APPLICATION NUMBER: US/10/264,283  
CURRENT FILING DATE: 2002-10-02  
NUMBER OF SEQ ID NOS: 111  
SOFTWARE: Corixa Invention Disclosure Database  
SEQ ID NO 90  
LENGTH: 253  
TYPE: PRT  
ORGANISM: Homo sapiens

US-10-264-283-90

Query Match 100.0%; Score 40; DB 14; Length 253;  
Best Local Similarity 100.0%; Pred. No. 18;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 LLPLQILL 9  
6 LLPLQILL 14

RESULT 6

US-10-264-283-90  
Sequence 90, Application US/10264283  
Publication No. US2003014494A1

GENERAL INFORMATION:

APPLICANT: Algate, Paul A.  
APPLICANT: Mannion, Jane  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
FILE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER  
FILE REFERENCE: 210121.590  
CURRENT APPLICATION NUMBER: US/10/264,283  
CURRENT FILING DATE: 2002-10-02  
NUMBER OF SEQ ID NOS: 111  
SOFTWARE: Corixa Invention Disclosure Database  
SEQ ID NO 90  
LENGTH: 253  
TYPE: PRT  
ORGANISM: Homo sapiens

US-10-264-283-90

Query Match 100.0%; Score 40; DB 14; Length 253;  
Best Local Similarity 100.0%; Pred. No. 18;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 LLPLQILL 9  
6 LLPLQILL 14

## RESULT 6

3-10-295-027-498  
Sequence 498, Application US/10295027  
Publication No. US2003023350A1

## GENERAL INFORMATION:

APPLICANT: Afar, Daniel  
APPLICANT: Aziz, Natasha  
APPLICANT: Ginsberg, Wendy M.  
APPLICANT: Gish, Kurt C.  
APPLICANT: Glynn, Richard  
APPLICANT: Hevezi, Peter A.  
APPLICANT: Mack, David H.  
APPLICANT: Murray, Richard  
APPLICANT: Watson, Susan R.

APPLICANT: Eos Biotechnology, Inc.

TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and  
TITLE OF INVENTION: Methods of Screening for Modulators of Cancer

FILE REFERENCE: 018501-012500US

CURRENT APPLICATION NUMBER: US/10/295,027

PRIOR FILING DATE: 2002-11-13

PRIOR APPLICATION NUMBER: US 09/663,733

PRIOR FILING DATE: 2000-09-15

PRIOR APPLICATION NUMBER: US 60/350,666

PRIOR FILING DATE: 2001-11-13

PRIOR APPLICATION NUMBER: US 60/335,394

PRIOR FILING DATE: 2001-11-15

PRIOR APPLICATION NUMBER: US 60/332,464

PRIOR FILING DATE: 2001-11-21

PRIOR APPLICATION NUMBER: US 60/334,393

PRIOR FILING DATE: 2001-11-29

PRIOR APPLICATION NUMBER: US 60/340,376

PRIOR FILING DATE: 2001-12-14

PRIOR APPLICATION NUMBER: US 60/347,211

PRIOR FILING DATE: 2002-01-08

PRIOR APPLICATION NUMBER: US 60/347,349

PRIOR FILING DATE: 2002-01-10

PRIOR APPLICATION NUMBER: US 60/355,250

PRIOR FILING DATE: 2002-02-08

PRIOR APPLICATION NUMBER: US 60/356,714

PRIOR FILING DATE: 2002-02-13

Remaining Prior Application data removed - See File Wrapper or PALM.

NUMBER OF SEQ ID NOS: 1386

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 498

LENGTH: 253

TYPE: PRT

ORGANISM: Homo sapiens

S-10-295-027-498

Query Match 100.0%; Score 40; DB 15; Length 253;  
Best Local Similarity 100.0%; Pred. No. 18;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 LLPLQILL 9  
b 6 LLPLQILL 14

## RESULT 7

S-10-173-999-48

Sequence 48, Application US/10173999

Publication No. US20040005563A1

## GENERAL INFORMATION:

APPLICANT: Mack, David H.

APPLICANT: Gish, Kurt C.

APPLICANT: Eos Biotechnology, Inc.

TITLE OF INVENTION: Methods of Diagnosis of Ovarian Cancer, Compositions

TITLE OF INVENTION: and Methods of Screening for Modulators of Ovarian

TITLE OF INVENTION: Cancer

FILE REFERENCE: 018501-002420US

CURRENT APPLICATION NUMBER: US/10/173,999

PRIOR FILING DATE: 2002-06-17

PRIOR APPLICATION NUMBER: US 60/299,234

PRIOR APPLICATION NUMBER: US 60/299,234

PRIOR APPLICATION NUMBER: US 60/299,234

PRIOR FILING DATE: 2001-06-18  
PRIOR APPLICATION NUMBER: US 60/315,287  
PRIOR FILING DATE: 2001-08-27  
PRIOR APPLICATION NUMBER: US 60/350,666  
PRIOR FILING DATE: 2001-11-13  
PRIOR APPLICATION NUMBER: US 60/372,246  
PRIOR FILING DATE: 2001-04-12  
NUMBER OF SEQ ID NOS: 163  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 48  
LENGTH: 253  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-173-999-48

Query Match 100.0%; Score 40; DB 15; Length 253;  
Best Local Similarity 100.0%; Pred. No. 18;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLPLQILL 9  
Db 6 LLPLQILL 14

## RESULT 8

US-09-918-243-33

Sequence 33, Application US/09918243

Patent No. US20020142317A1

GENERAL INFORMATION:

APPLICANT: O'Brien, Timothy J.

APPLICANT: Cannon, Martin J.

APPLICANT: Santin, Alessandro

TITLE OF INVENTION: Methods for the early diagnosis of ovarian cancer

FILE REFERENCE: D6223CIP/C/D/CIP

CURRENT APPLICATION NUMBER: US/09/918,243

CURRENT FILING DATE: 2001-07-30

PRIOR APPLICATION NUMBER: US

PRIOR FILING DATE: 2001-07-13

NUMBER OF SEQ ID NOS: 136

SEQ ID NO 33

LENGTH: 9

TYPE: PRT

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: CHAIN

OTHER INFORMATION: Residues 5-13 of the SCCE protein

US-09-918-243-33

Query Match 90.0%; Score 36; DB 9; Length 9;

Best Local Similarity 100.0%; Pred. No. 7.1e+05;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLPLQILL 8  
Db 2 LLPLQILL 9

## RESULT 9

US-09-905-083-33

Sequence 33, Application US/09905083

Patent No. US20020146708A1

GENERAL INFORMATION:

APPLICANT: O'Brien, Timothy J.

TITLE OF INVENTION: Compositions and Methods for the Early Diagnosis of

TITLE OF INVENTION: Ovarian Cancer

FILE REFERENCE: D6223CIP/C/DiV

CURRENT APPLICATION NUMBER: US/09/905,083

CURRENT FILING DATE: 2001-07-13

PRIOR APPLICATION NUMBER: US 09/502,600

PRIOR FILING DATE: 2000-02-11

NUMBER OF SEQ ID NOS: 136

SEQ ID NO 33

LENGTH: 9

```

TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CHAIN
OTHER INFORMATION: Residues 5-13 of the SCCE protein
S-09-905-083-33

Query Match          90.0%; Score 36; DB 9; Length 9;
Best Local Similarity 100.0%; Pred. No. 7.1e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

y 1 LLPLQILL 8
b 2 LLPLQILL 9

RESULT 10
S-10-369-493-10255
Sequence 10255, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 10255
LENGTH: 868
TYPE: PRT
ORGANISM: Cytophaga hutchinsonii
S-10-369-493-10255

Query Match          90.0%; Score 36; DB 15; Length 868;
Best Local Similarity 88.9%; Pred. No. 3.2e+02;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

y 1 LLPLQILL 9
b 704 LLPLQILL 712

RESULT 11
S-09-864-761-46097
Sequence 46097, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
FILE REFERENCE: Aeonica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27

```

```

; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 46097
; LENGTH: 23
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AJ246003.1
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.6
US-09-864-761-46097

Query Match          87.5%; Score 35; DB 9; Length 23;
Best Local Similarity 77.8%; Pred. No. 12;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

y 1 LLPLQILL 9
b 11 LLPLQILL 19

RESULT 12
US-09-895-913A-92
Sequence 92, Application US/09895913A
Patent No. US20020160456A1
GENERAL INFORMATION:
APPLICANT: Kleanthous, Harold
APPLICANT: Al-Garawi, Amal
APPLICANT: Miller, Charles
APPLICANT: Tomb, Jean Francois
APPLICANT: Omen, Raymond P.
TITLE OF INVENTION: Identification of Polynucleotides
TITLE OF INVENTION: Encoding No. US20020160456A1 Helicobacter Polypeptides in the
TITLE OF INVENTION: Genome
FILE REFERENCE: 06132/043002
CURRENT APPLICATION NUMBER: US/09/895,913A
CURRENT FILING DATE: 2001-06-29
PRIOR APPLICATION NUMBER: US 08/881,227
PRIOR FILING DATE: 1997-06-24
NUMBER OF SEQ ID NOS: 368
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 92
LENGTH: 480
TYPE: PRT

```

ORGANISM: Helicobacter pylori  
S-09-895-913A-92

Query Match 87.5%; Score 35; DB 9; Length 480;  
Best Local Similarity 88.9%; Pred. No. 2.6e+02;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Y 1 LPLQLILL 9  
||| |||  
b 238 LPLQLILL 246

RESULT 13  
S-10-097-065-220  
Sequence 220, Application US/10097065  
Publication No. US20030055236A1  
GENERAL INFORMATION:  
APPLICANT: Moore, Paul A. et al.  
TITLE OF INVENTION: 110 Human Secreted Proteins  
FILE REFERENCE: P2021P1  
CURRENT APPLICATION NUMBER: US/10/097,065  
PRIOR APPLICATION NUMBER: PCT/US98/27059  
PRIOR FILING DATE: 1998-12-17  
PRIOR APPLICATION NUMBER: 60/070,923  
PRIOR FILING DATE: 1997-12-18  
PRIOR APPLICATION NUMBER: 60/068,007  
PRIOR FILING DATE: 1997-12-18  
PRIOR APPLICATION NUMBER: 60/068,057  
PRIOR FILING DATE: 1997-12-18  
PRIOR APPLICATION NUMBER: 60/068,006  
PRIOR FILING DATE: 1997-12-18  
PRIOR APPLICATION NUMBER: 60/068,369  
PRIOR FILING DATE: 1997-12-19  
PRIOR APPLICATION NUMBER: 60/068,367  
PRIOR FILING DATE: 1997-12-19  
PRIOR APPLICATION NUMBER: 60/068,368  
PRIOR FILING DATE: 1997-12-19  
PRIOR APPLICATION NUMBER: 60/068,169  
PRIOR FILING DATE: 1997-12-19  
PRIOR APPLICATION NUMBER: 60/068,053  
PRIOR FILING DATE: 1997-12-18  
PRIOR APPLICATION NUMBER: 60/068,064  
PRIOR FILING DATE: 1997-12-18  
PRIOR APPLICATION NUMBER: 60/068,054  
PRIOR FILING DATE: 1997-12-18  
PRIOR APPLICATION NUMBER: 60/068,008  
PRIOR FILING DATE: 1997-12-18  
PRIOR APPLICATION NUMBER: 60/068,365  
PRIOR FILING DATE: 1997-12-19  
NUMBER OF SEQ ID NOS: 672  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 220  
LENGTH: 70  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (70)  
OTHER INFORMATION: Xaa equals stop translation  
US-10-097-065-220

Query Match 85.0%; Score 34; DB 14; Length 70;  
Best Local Similarity 87.5%; Pred. No. 56;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 LPLQLILL 9  
||| |||  
Db 5 LPLQLILL 12

RESULT 14  
US-09-933-767-1083

; Sequence 1083, Application US/09933767  
; Publication No. US20030181692A1  
; GENERAL INFORMATION:  
; APPLICANT: Ni et al.  
; TITLE OF INVENTION: 207 Human Secreted Proteins  
; FILE REFERENCE: P2007P2  
; CURRENT APPLICATION NUMBER: US/09/933,767  
; CURRENT FILING DATE: 2001-08-22  
; PRIOR APPLICATION NUMBER: PCT/US01/05614  
; PRIOR FILING DATE: 2001-02-21  
; PRIOR APPLICATION NUMBER: 60/184,836  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: 60/193,170  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: 09/205,258  
; PRIOR FILING DATE: 1998-12-04  
; PRIOR APPLICATION NUMBER: PCT/US98/11422  
; PRIOR FILING DATE: 1998-06-04  
; PRIOR APPLICATION NUMBER: 60/048,885  
; PRIOR FILING DATE: 1997-06-06  
; PRIOR APPLICATION NUMBER: 60/049,375  
; PRIOR FILING DATE: 1997-06-06  
; PRIOR APPLICATION NUMBER: 60/048,881  
; PRIOR FILING DATE: 1997-06-06  
; PRIOR APPLICATION NUMBER: 60/048,880  
; PRIOR FILING DATE: 1997-06-06  
; PRIOR APPLICATION NUMBER: 60/048,896  
; PRIOR FILING DATE: 1997-06-06  
; PRIOR APPLICATION NUMBER: 60/049,020  
; PRIOR FILING DATE: 1997-06-06  
; PRIOR APPLICATION NUMBER: 60/048,876  
; PRIOR FILING DATE: 1997-06-06  
; PRIOR APPLICATION NUMBER: 60/048,895  
; PRIOR FILING DATE: 1997-06-06  
; PRIOR APPLICATION NUMBER: 60/048,884  
; PRIOR FILING DATE: 1997-06-06  
; PRIOR APPLICATION NUMBER: 60/048,894  
; PRIOR FILING DATE: 1997-06-06  
; PRIOR APPLICATION NUMBER: 60/048,971  
; PRIOR FILING DATE: 1997-06-06  
; PRIOR APPLICATION NUMBER: 60/048,964  
; PRIOR FILING DATE: 1997-06-06  
; PRIOR APPLICATION NUMBER: 60/048,882  
; PRIOR FILING DATE: 1997-06-06  
; PRIOR APPLICATION NUMBER: 60/048,899  
; PRIOR FILING DATE: 1997-06-06  
; PRIOR APPLICATION NUMBER: 60/048,893  
; PRIOR FILING DATE: 1997-06-06  
; PRIOR APPLICATION NUMBER: 60/048,900  
; PRIOR FILING DATE: 1997-06-06  
; PRIOR APPLICATION NUMBER: 60/048,901  
; PRIOR FILING DATE: 1997-06-06  
; PRIOR APPLICATION NUMBER: 60/048,892  
; PRIOR FILING DATE: 1997-06-06  
; PRIOR APPLICATION NUMBER: 60/048,915  
; PRIOR FILING DATE: 1997-06-06  
; PRIOR APPLICATION NUMBER: 60/049,019  
; PRIOR FILING DATE: 1997-06-06  
; PRIOR APPLICATION NUMBER: 60/048,970  
; PRIOR FILING DATE: 1997-06-06  
; PRIOR APPLICATION NUMBER: 60/048,972  
; PRIOR FILING DATE: 1997-06-06  
; PRIOR APPLICATION NUMBER: 60/048,916  
; PRIOR FILING DATE: 1997-06-06  
; PRIOR APPLICATION NUMBER: 60/049,373  
; PRIOR FILING DATE: 1997-06-06  
; PRIOR APPLICATION NUMBER: 60/048,875  
; PRIOR FILING DATE: 1997-06-06  
; PRIOR APPLICATION NUMBER: 60/049,374  
; PRIOR FILING DATE: 1997-06-06  
; PRIOR APPLICATION NUMBER: 60/048,917  
; PRIOR FILING DATE: 1997-06-06  
; PRIOR APPLICATION NUMBER: 60/048,949

PRIOR FILING DATE: 1997-06-06  
PRIOR APPLICATION NUMBER: 60/048,974  
PRIOR FILING DATE: 1997-06-06  
PRIOR APPLICATION NUMBER: 60/048,883  
PRIOR FILING DATE: 1997-06-06  
PRIOR APPLICATION NUMBER: 60/048,897  
PRIOR FILING DATE: 1997-06-06  
PRIOR APPLICATION NUMBER: 60/048,898  
PRIOR FILING DATE: 1997-06-06  
PRIOR APPLICATION NUMBER: 60/048,962  
PRIOR FILING DATE: 1997-06-06  
PRIOR APPLICATION NUMBER: 60/048,963  
PRIOR FILING DATE: 1997-06-06  
PRIOR APPLICATION NUMBER: 60/048,877  
PRIOR FILING DATE: 1997-06-06  
PRIOR APPLICATION NUMBER: 60/048,878  
PRIOR FILING DATE: 1997-06-06  
PRIOR APPLICATION NUMBER: 60/068,054  
PRIOR FILING DATE: 1997-12-18  
PRIOR APPLICATION NUMBER: 60/068,064  
PRIOR FILING DATE: 1997-12-18  
PRIOR APPLICATION NUMBER: 60/068,053  
PRIOR FILING DATE: 1997-12-18  
PRIOR APPLICATION NUMBER: 60/070,923  
PRIOR FILING DATE: 1997-12-18  
PRIOR APPLICATION NUMBER: 60/073,160  
PRIOR FILING DATE: 1998-01-30  
PRIOR APPLICATION NUMBER: 60/073,159  
PRIOR FILING DATE: 1998-01-30  
PRIOR APPLICATION NUMBER: 60/073,165  
PRIOR FILING DATE: 1998-01-30  
PRIOR APPLICATION NUMBER: 60/073,164  
PRIOR FILING DATE: 1998-01-30  
PRIOR APPLICATION NUMBER: 60/085,925  
PRIOR FILING DATE: 1998-05-18  
PRIOR APPLICATION NUMBER: 60/085,921  
PRIOR FILING DATE: 1998-05-18  
PRIOR APPLICATION NUMBER: 60/085,923  
PRIOR FILING DATE: 1998-05-18  
PRIOR APPLICATION NUMBER: 60/085,922  
PRIOR FILING DATE: 1998-05-18  
PRIOR APPLICATION NUMBER: 60/092,921  
PRIOR FILING DATE: 1998-07-15  
PRIOR APPLICATION NUMBER: 60/094,657  
PRIOR FILING DATE: 1998-07-30  
NUMBER OF SEQ ID NOS: 1245  
SOFTWARE: Patent in Ver. 2.0  
SEQ ID NO 1083  
LENGTH: 73  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (8)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
S-09-933-767-1083  
Query Match 85.0%; Score 34; DB 10; Length 73;  
Best Local Similarity 87.5%; Pred. No. 59;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
y 2 LPLQLLL 9  
|||:|  
b 25 LPLQLLL 32  
RESULT 15  
S-10-023-1083  
Sequence 1083, Application US/10023282  
Publication No. US2003092893A1  
GENERAL INFORMATION:  
APPLICANT: Young et al.  
TITLE OF INVENTION: 207 Human Secreted Proteins  
FILE REFERENCE: PZ007P1  
CURRENT APPLICATION NUMBER: US/10/023,282  
CURRENT FILING DATE: 2001-12-20  
EARLIER APPLICATION NUMBER: 09/205,258  
EARLIER FILING DATE: 1998-12-04  
EARLIER APPLICATION NUMBER: PCT/US98/11422  
EARLIER FILING DATE: 1998-06-04  
EARLIER APPLICATION NUMBER: 60/048,885  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/049,375  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,881  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,880  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,896  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/049,020  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,876  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,895  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,884  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,894  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,971  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,964  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,882  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,899  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,893  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,900  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,901  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,892  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,915  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/049,019  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,970  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,972  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,916  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/049,373  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,875  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/049,374  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,917  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,949  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,974  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,883  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,897  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,898  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,962  
EARLIER FILING DATE: 1997-06-06

EARLIER APPLICATION NUMBER: 60/048,963  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,877  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,878  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/070,923  
EARLIER FILING DATE: 1997-12-18  
EARLIER APPLICATION NUMBER: 60/092,921  
EARLIER FILING DATE: 1998-07-15  
EARLIER APPLICATION NUMBER: 60/094,657  
EARLIER FILING DATE: 1998-07-30  
NUMBER OF SEQ ID NOS: 1227  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 1083  
LENGTH: 73  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (8)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
S-10-023-282-1083  
Query Match 85.0%; Score 34; DB 14; Length 73;  
Best Local Similarity 87.5%; Pred. No. 59;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
Y 2 LPLQILL 9  
b 25 LPLQILL 32  
Search completed: March 1, 2004, 18:08:51  
Job time : 25.1111 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

protein - protein search, using sw model

on: March 1, 2004, 17:21:01 ; Search time 10.2222 Seconds  
(without alignments)  
84.690 Million cell updates/sec

title: US-09-905-083-35

effect score: 40

sequence: 1 LLPLQLILL 9

coring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

searched: 283366 seqs, 96191526 residues

total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 78:\*

1: PIR1:\*

2: PIR2:\*

3: PIR3:\*

4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	40	100.0	253	A53968	serine proteinase
2	36	90.0	845	D97163	cation transport P
3	35	87.5	493	A71875	hypothetical prote
4	35	87.5	496	B64638	conserved hypothet
5	34	85.0	266	JC7300	tax-responsive ele
6	34	85.0	267	JC4857	hepatocarcinogenes
7	34	85.0	389	S33603	surfactant protein
8	34	85.0	371	JN0450	conglutinin precu
9	34	85.0	371	I45878	conglutinin - bovi
10	34	85.0	429	T17215	hypothetical prote
11	34	85.0	539	H82994	probable permease
12	34	85.0	1092	T12520	hypothetical prote
13	34	85.0	1236	H50904	Mg protoporphyrin
14	33	82.5	146	T50201	hypothetical prote
15	33	82.5	277	H84314	cytochrome a3 con
16	33	82.5	282	E70890	hypothetical prote
17	33	82.5	294	A25970	transcription acti
18	33	82.5	294	E82257	cholera toxin tran
19	33	82.5	653	A46362	amyloid precursor-
20	32	80.0	245	E75440	hypothetical prote
21	32	80.0	253	AF1535	transport protein
22	32	80.0	256	JQ0708	lipopolysaccharide
23	32	80.0	394	I B43750	chloramphenicol re
24	32	80.0	394	B85535	probable transport
25	32	80.0	394	F90684	probable transport
26	32	80.0	412	A57468	P-selectin glycopo
27	32	80.0	491	JC6197	stromelysin 3 (EC
28	32	80.0	492	A44399	stromelysin 3 (EC
29	32	80.0	543	A12088	Na+/H+-exchanging

## ALIGNMENTS

### RESULT 1

A53968  
serine proteinase SCCE precursor - human  
N:Alternate names: stratum corneum chymotryptic enzyme  
C:Species: Homo sapiens (man)  
C:Date: 07-Jul-1995 #sequence\_revision 07-Jul-1995 #text\_change 22-Jun-1999  
C:Accession: A53968  
R:Hansson, L.; Siroenqvist, M.; Baeckman, A.; Wallbrandt, P.; Carlstein, A.; Egelrud, T.  
J. Biol. Chem. 269, 19420-19426, 1994  
A:Title: Cloning, expression, and characterization of stratum corneum chymotryptic enzym  
A:Reference number: A53968; MUID:94308225; PMID:8034709  
A:Accession: A53968  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-253 <HAN>  
A:Cross-references: GB:L33404; NID:G521214; PIDN:AAC37551.1; PID:G532504  
C:Genetics:  
A:Gene: GDB:PS56; SCCE  
A:Cross-references: GDB:377730  
A:Map position: 7q35-7q35  
A:Superfamily: trypsin; trypsin homology  
F:30-245/Domain: trypsin homology <TRY>

Query Match 100.0%; Score 40; DB 2; Length 253;  
Best Local Similarity 100.0%; Pred. No. 1.9;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 LLPLQLILL 9  
DB 6 LLPLQLILL 14

### RESULT 2

D97163  
cation transport P-type ATPase CAC2137 [imported] - Clostridium acetobutylicum  
C:Species: Clostridium acetobutylicum  
C:Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 14-Sep-2001  
C:Accession: D97163  
R:Nolling, J.; Bretton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,  
J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
J. Bacteriol. 183, 4823-4838, 2001  
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo  
A:Reference number: A96900; MUID:21359325; PMID:21359325  
A:Accession: D97163  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-845 <KUR>  
A:Cross-references: GB:AB001437; PIDN:AAK80095.1; PID:G15025128; GSPDB:GN00168  
A:Experimental source: Clostridium acetobutylicum ATCC824  
C:Genetics:  
A:Gene: CAC2137

30 32 80.0 1041 2 T29010  
31 32 80.0 1306 1 A31759  
32 32 80.0 1485 2 S21756  
33 31 77.5 157 2 A83214  
34 31 77.5 189 2 T43786  
35 31 77.5 196 2 G65039  
36 31 77.5 209 2 A54984  
37 31 77.5 213 2 JE0322  
38 31 77.5 303 2 C84914  
39 31 77.5 306 2 H86811  
40 31 77.5 327 1 A41720  
41 31 77.5 328 1 A41927  
42 31 77.5 370 2 AB3334  
43 31 77.5 372 2 A98157  
44 31 77.5 372 2 AH3130  
45 31 77.5 388 2 T26360

hypothetical prote  
peptidyl-dipeptida  
CFTR protein - Afr  
probable transcrip  
hypothetical prote  
hypothetical prote  
ELF-1 protein prec  
ephraim-A2 - human  
hypothetical prote  
sugar ABC transpor  
acid phosphatase (  
insulin-like growt  
daunorubicin resis  
probable permease  
ABC transporter, m  
hypothetical prote

```

Query Match          90.0%; Score 36; DB 2; Length 845;
Best Local Similarity 77.8%; Pred. No. 39;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

1 LLPQLILL 9
|||:||||:
670 LLPQLILL 678

SULT 3
1875
putational protein jhp0890 - Helicobacter pylori (strain J99)
Species: Helicobacter pylori
Variety: strain J99
Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 08-Oct-1999
Accession: A71875
Alm, R.A.; King, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;
Ives, C.; Gibson, R.; Werberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;
ature 397, 176-180, 1999
Title: Genomic sequence comparison of two unrelated isolates of the human gastric path
Reference number: A71800; MUID:99120557; PMID:9923682
Accession: A71875
Status: preliminary
Molecule type: DNA
Residues: 1-493 <ARN>
Cross-references: GB:AE001518; GB:AE001439; NID:94155454; PIDN:AA006464.1; PID:9415546
Experimental source: strain J99
Genetics:
Gene: jhp0880

Query Match          87.5%; Score 35; DB 2; Length 493;
Best Local Similarity 88.9%; Pred. No. 36;
Matches 8; Conservative 0; Mismatches 1; Indels 1; Gaps 0;

1 LLPQLILL 9
|||:||||:
254 LLPQLILL 262

SULT 4
54638
nserved hypothetical integral membrane protein HP0946 - Helicobacter pylori (strain 28
Species: Helicobacter pylori
Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 08-Oct-1999
Accession: B64638
Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.;
Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKen
on, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.
ature 388, 539-547, 1997
Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
Reference number: A64520; MUID:97394467; PMID:9252185
Accession: B64638
Status: preliminary; nucleic acid sequence not shown; translation not shown
Molecule type: DNA
Residues: 1-496 <ROM>
Cross-references: GB:AE000604; GB:AE000511; NID:92314086; PIDN:AA007993.1; PID:9231408
Genetics:
Start codon: GTG

Query Match          87.5%; Score 35; DB 2; Length 496;
Best Local Similarity 88.9%; Pred. No. 36;
Matches 8; Conservative 0; Mismatches 1; Indels 1; Gaps 0;

1 LLPQLILL 9
|||:||||:
254 LLPQLILL 262

SULT 5
C7300
ax-responsive element-binding protein 5 - mouse
Species: Mus musculus (house mouse)

```

---

```

C>Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 02-Sep-2000
C/Accession: JC7300
R/Masaki, T.; Noguchi, H.; Kobayashi, M.; Yoshida, M.; Takamatsu, K.
DNA Res. 7, 187-193, 2000
A/Title: Isolation and characterization of the gene encoding mouse tax-responsive element
A/Reference number: JC7300
A/Accession: JC7300
A/Molecule type: mRNA
A/Residues: 1-266 <MAS>
A/Cross-references: DDBJ:AB036745
C/Genetics:
A/Gene: treb5
C/Superfamily: X box-binding protein 1; fos/jun DNA-binding domain homology
C/Keywords: leucine zipper; transcription factor

Query Match          85.0%; Score 34; DB 2; Length 266;
Best Local Similarity 88.9%; Pred. No. 30;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLPQLILL 9
|||:||||:
187 LLPQLILL 195

Db

RESULT 6
JC4857
hepatocarcinogenesis-related transcription factor - rat
C/Species: Rattus norvegicus (Norway rat)
C/Date: 15-Aug-1996 #sequence_revision 18-Oct-1996 #text_change 08-Sep-1997
C/Accession: JC4857
R/Kishimoto, T.; Kokura, K.; Kumagai, Y.; Makino, Y.; Tamura, T.
Biochem. Biophys. Res. Commun. 224, 746-751, 1996
A/Title: HTF: A b-zip transcription factor that is closely related to the human XBP/TREB:
A/Reference number: JC4857
A/Accession: JC4857
A/Molecule type: mRNA
A/Residues: 1-267 <KIS>
C/Comment: This is a basic-leucine zipper type transcription factor involved in hepatocel
C/Genetics:
A/Gene: htF
C/Superfamily: X box-binding protein 1; fos/jun DNA-binding domain homology
C/Keywords: leucine zipper; transcription factor
F/89-98/Domain: fos/jun DNA-binding domain homology <FUD>
F/89-126/Region: leucine zipper motif

Query Match          85.0%; Score 34; DB 2; Length 267;
Best Local Similarity 88.9%; Pred. No. 30;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLPQLILL 9
|||:||||:
188 LLPQLILL 196

Db

RESULT 7
S33603
surfactant protein D - bovine
C/Species: Bos primigenius taurus (cattle)
C/Date: 02-Dec-1993 #sequence_revision 01-Sep-1995 #text_change 17-Mar-1999
C/Accession: S33603
R/Lim, B.L.; Lu, J.; Reid, K.B.M.
Immunology 78, 159-165, 1993
A/Title: Structural similarity between bovine conglutinin and bovine lung surfactant prot
A/Reference number: S33603; MUID:93170856; PMID:8436402
A/Accession: S33603
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-369 <IRM>
C/Superfamily: pulmonary surfactant protein D; C-type lectin homology
F/248-367/Domain: C-type lectin homology <LCH>

Query Match          85.0%; Score 34; DB 2; Length 369;
Best Local Similarity 87.8%; Pred. No. 42;

```



```

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
1 LLPLQIILL 9
  ||||:||||
3 LLPLSVLL 11

:ISULT 8
:ID450
nglutinin precursor - bovine
Alternate names: C3b-binding protein
Contains: conglutinin-N
Species: Bos primigenius taurus (cattle)
Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
Accession: UN0450; JG2396; S3235; A23740; S35044; I46010; A29416; S34054
Suzuki, Y.; Yin, Y.P.; Makino, M.; Kurimura, T.; Wakamiya, N.
ochem. Biophys. Res. Commun. 191, 335-342, 1993
Title: Cloning and sequencing of a cDNA coding for bovine conglutinin.
Reference number: UN0450; MUID:93213261; PMID:8460993
Accession: UN0450
Molecule type: mRNA
Residues: 1-371 <SUZ>
Cross-references: DBJ:U14085; NID:G285643; PIDN:BAA03170.1; PID:G285644
Experimental source: liver
Kawasaki, N.; Itoh, N.; Kawasaki, T.
ochem. Biophys. Res. Commun. 198, 597-604, 1994
Title: Gene organization and 5'-flanking region sequence of conglutinin: A C-type mamm
Reference number: JG2396; MUID:94128104; PMID:8297370
Accession: JG2396
Molecule type: mRNA
Residues: 1-371 <KAZ>
Note: The authors translated the codon CAT for residues 250 and 270 as Glu
Lu, J.; Laursen, S.B.; Thiel, S.; Jensenius, J.C.; Reid, K.B.M.
ochem. J. 292, 157-162, 1993
Title: The cDNA cloning of conglutinin and identification of liver as a primary site o
Reference number: S3235; MUID:93277452; PMID:7684896
Accession: S3235
Molecule type: mRNA
Residues: 1-172, 'H', 174-217, 'A', 219-271, 'V', 273-371 <LUU>
Cross-references: EMBL:X71774; NID:G395267; PIDN:CAA50665.1; PID:G395268
Experimental source: liver
Lee, Y.M.; Leiby, K.R.; Aliar, J.; Paris, K.; Lerch, B.; Okarma, T.B.
Biol. Chem. 266, 2715-2723, 1991
Title: Primary structure of bovine conglutinin, a member of the C-type animal lectin f
Reference number: A23740; MUID:91131556; PMID:1993651
Accession: A23740
Molecule type: protein
Residues: 21-209, 'S', 211-371 <LEE>
Kawasaki, N.; Yokota, Y.; Kawasaki, T.
rch. Biochem. Biophys. 305, 533-540, 1993
Title: Differentiation of conglutination activity and sugar-binding activity of conglu
Reference number: S36879; MUID:93384312; PMID:8373191
Accession: S36879
Molecule type: protein
Residues: 21-54; 75-86, 'X', 88-89, 'X', 91, 'X', 93-94; 208-209, 'X', 211-227 <KAN>
Experimental source: serum
Lu, J.; Wiedemann, H.; Holmikov, U.; Thiel, S.; Timpl, R.; Reid, K.B.M.
ur. J. Biochem. 215, 793-799, 1993
Title: Structural similarity between lung surfactant protein D and conglutinin. Two di
Reference number: S35044; MUID:93358905; PMID:8354286
Accession: S35044
Molecule type: protein
Residues: 75-86, 'X', 88-89, 'X', 91, 'I' <LUN>
Experimental source: lung
Young, N.M.; Leon, M.A.
ochem. Biophys. Res. Commun. 143, 645-651, 1987
Title: The carbohydrate specificity of conglutinin and its homology to proteins in the
Reference number: A29416; MUID:87184551; PMID:3566740
Accession: A29416
Contents: annotation
Walhotra, R.; Laursen, S.B.; Willis, A.C.; Sim, R.B.
ochem. J. 293, 15-19, 1993
Title: Research Communication. Localization of the receptor-binding site in the collec
Reference number: S34054; MUID:93319501; PMID:8328957
Accession: S34054

A;Contents: annotation
R;Liou, L.S.; Sastry, R.; Hartshorn, K.L.; Lee, Y.M.; Okarma, T.B.; Tauber, A.I.; Sastry,
J. Immunol. 153, 173-180, 1994
A;Title: Bovine conglutinin gene exon structure reveals its evolutionary relationship to
A;Reference number: I46010; MUID:94267222; PMID:8207234
A;Accession: I46010
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-371 <LIO>
A;Cross-references: EMBL:U06860; NID:G507183; PIDN:AAB0624.1; PID:G514256
C;Comment: This protein mediates the agglutination of erythrocytes with antibody and comp
C;Comment: This protein is a Ca2+-dependent serum lectin specific for N-acetylglucosamine
C;Genetics:
A;Gene: CNL
A;Introns: 67/1; 106/1; 142/1; 178/1; 217/1; 245/1
C;Superfamily: pulmonary surfactant protein D; C-type lectin homology
C;Keywords: calcium binding; glycoprotein; hydroxyllysine; hydroxyproline
F;1-20/Domain: signal sequence #status predicted <SIG>
F;21-371/Product: conglutinin #status predicted <MAT>
F;46-214/Region: collagen-like
F;45-371/Product: conglutinin-N #status predicted <MA2>
F;248-369/Domain: C-type lectin homology <LCH>
F;63,87,99,135,141,159,162,198,210/Binding site: carbohydrate (Lys) (covalent) #status p
F;63,87,99,135,141,159,162,198,210/Modified site: 5-hydroxyllysine (Lys) #status experim
F;78,96,108,111,129,132,147,153,171,195/Modified site: 4-hydroxyproline (Pro) #status exi
F;337/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 85.0%; Score 34; DB 1; Length 371;
Best Local Similarity 77.8%; Pred. No. 42;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LLPLQIILL 9
Db 3 LLPLSVLL 11

RESULT 9
I45878
conglutinin - bovine
C;Species: Bos primigenius taurus (cattle)
C;Date: 19-Dec-1997 #sequence_revision 19-Dec-1997 #text_change 20-Aug-1999
C;Accession: I45878
R;Liou, L.S.; Sastry, R.; Hartshorn, K.L.; Lee, Y.M.; Okarma, T.B.; Tauber, A.I.; Sastry,
Gene 141, 277-281, 1994
A;Title: Bovine conglutinin (BC) mRNA expressed in liver: cloning and characterization of
A;Reference number: I45878; MUID:94215917; PMID:8163202
A;Accession: I45878
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-371 <LIO>
A;Cross-references: GB:I18971; NID:G495012; PIDN:AAA20126.1; PID:G495013
C;Superfamily: pulmonary surfactant protein D; C-type lectin homology
F;248-369/Domain: C-type lectin homology <LCH>

Query Match 85.0%; Score 34; DB 2; Length 371;
Best Local Similarity 77.8%; Pred. No. 42;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LLPLQIILL 9
Db 3 LLPLSVLL 11

RESULT 10
T17215
hypothetical protein DKFZp434H2235.1 - human
C;Species: Homo sapiens (man)
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T17215
R;Poustka, A.; Klein, M.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, September 1999
A;Reference number: Z18723
A;Accession: T17215

```

;Status: preliminary  
;Molecule type: mRNA  
;Residues: 1-429 <POU>  
;Cross-references: EMBL:AL117404  
;Experimental source: adult testis; clone DKFZp434H2235  
;Genetics:  
;Note: DKFZp434H2235.1

Query Match 85.0%; Score 34; DB 2; Length 429;  
Best Local Similarity 87.5%; Pred. No. 49;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Y 2 LPLQLLLL 9  
|||:|  
b 381 LPLQLLLL 389

RESULT 11  
82994  
;Title: Probable permease of ABC iron transporter PA5216 [imported] - Pseudomonas aeruginosa (strain ATCC 27802)  
;Species: Pseudomonas aeruginosa  
;Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
;Accession: H82994  
;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Boman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, J.; Lory, S.; Olson, M.V.  
;ature 406, 959-964, 2000  
;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen  
;Reference number: A82950; MUID:20437337; PMID:10984043  
;Accession: H82994  
;Status: preliminary  
;Molecule type: DNA  
;Residues: 1-539 <STO>  
;Cross-references: GB:AE004934; GB:AE004091; NID:G9951515; PIDN:AAG08601.1; GSPDB:GN001  
;Experimental source: strain PA01  
;Genetics:  
;Gene: PA5216  
;Superfamily: sfuB protein

Query Match 85.0%; Score 34; DB 2; Length 539;  
Best Local Similarity 77.8%; Pred. No. 62;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Y 1 LPLQLLLL 9  
|||:|  
b 21 LPLSVLLL 29

RESULT 12  
112520  
;Title: Hypothetical protein DKFZp434G173.1 - human  
;Species: Homo sapiens (man)  
;Date: 23-Jul-1999 #sequence\_revision 23-Jul-1999 #text\_change 23-Jul-1999  
;Accession: T12520  
;Wambutt, R.; Heubner, D.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.  
submitted to the Protein Sequence Database, June 1999  
;Reference number: Z17524  
;Accession: T12520  
;Status: preliminary  
;Molecule type: mRNA  
;Residues: 1-1092 <WAM>  
;Cross-references: EMBL:AL080133  
;Experimental source: adult testis; clone DKFZp434G173  
;Genetics:  
;Note: DKFZp434G173.1

Query Match 85.0%; Score 34; DB 2; Length 1092;  
Best Local Similarity 87.5%; Pred. No. 1.3e+02;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Y 2 LPLQLLLL 9  
|||:|  
b 1044 LPLQLLLL 1051

# RESULT 13

T50904  
;Title: Mg protoporphyrin methyl transferase [imported] - Rubrivivax gelatinosus  
;Species: Rubrivivax gelatinosus  
;Date: 21-Jul-2000 #sequence\_revision 21-Jul-2000 #text\_change 02-Sep-2000  
;Accession: T50904  
;Nagashima, K.V.; Igarashi, N.; Harada, J.; Nagashima, S.; Matsuura, K.; Shimada, K.  
submitted to the EMBL Data Library, November 1999  
;Description: Determination of Nucleotide Sequences of Rubrivivax gelatinosus Photosynthet  
;Reference number: Z25270  
;Accession: T50904  
;Status: preliminary; translated from GB/EMBL/DBJ  
;Molecule type: DNA  
;Residues: 1-1236 <NAG>  
;Cross-references: EMBL:AB034704; PIDN:BAA94057.1  
;Experimental source: strain IL144  
;Genetics:  
;Gene: bchH  
;Superfamily: Rhodobacter capsulatus magnesium-protoporphyrin O-methyltransferase

Query Match 85.0%; Score 34; DB 2; Length 1236;  
Best Local Similarity 88.9%; Pred. No. 1.4e+02;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Y 1 LLPLQLLLL 9  
|||:|  
b 946 LLPLQLKL 954

# RESULT 14

H75201  
;Title: Hypothetical protein PAB0088 - Pyrococcus abyssi (strain Orsay)  
;Species: Pyrococcus abyssi  
;Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 08-Sep-2000  
;Accession: H75201  
;Anonymous, Genoscope  
submitted to the EMBL Data Library, July 1999  
;Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome struc  
;Reference number: A75001  
;Accession: H75201  
;Status: preliminary  
;Molecule type: DNA  
;Residues: 1-146 <RAW>  
;Cross-references: GB:AJ248283; GB:AL096836; NID:G5457433; PIDN:CAB49063.1; PID:el51495;  
;Experimental source: strain Orsay  
;Genetics:  
;Gene: PAB0088  
;Superfamily: Pyrococcus abyssi hypothetical protein PAB0088

Query Match 82.5%; Score 33; DB 2; Length 146;  
Best Local Similarity 66.7%; Pred. No. 26;  
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Y 1 LLPLQLLLL 9  
|||:|  
b 68 LLPLQLIIV 76

# RESULT 15

H84314  
;Title: Cytochrome a3 controlling protein [imported] - Halobacterium sp. NRC-1  
;Species: Halobacterium sp. NRC-1  
;Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
;Accession: H84314  
;Rung, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.; Leitch, K.H.; Alam, M.; Freitas, T.  
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000  
;Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Lie  
;Title: Genome sequence of Halobacterium species NRC-1  
;Reference number: A84160; MUID:20504483; PMID:11016950

;Accession: H84314  
;Status: preliminary  
;Molecule type: DNA  
;Residues: 1-277 <STO>  
;Cross-references: GB:AE004437; NID:G10581096; PIDN:AG19884.1; GSPDB:GN00138  
;Genetics:  
;Gene: ccp

Query Match 82.5%; Score 33; DB 2; Length 277;  
Best Local Similarity 75.0%; Pred. No. 49;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Y 1 LFLQLLL 8  
| | | | |  
b 98 LFLQVIL 105

Search completed: March 1, 2004, 17:36:30  
Job time : 12.3333 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

#### 4 protein - protein search, using sw model

on: March 1, 2004, 17:17:25 ; Search time 6 seconds  
(without alignments)  
78.105 Million cell updates/sec

title: US-09-905-083-35

effect score: 40

sequence: 1 LLPQLILL 9

coring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

searched: 141681 seqs, 52070155 residues

total number of hits satisfying chosen parameters: 141681

minimum DB seq length: 0

maximum DB seq length: 2000000000

post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	40	100.0	253	1	KLK7_HUMAN
2	38	95.0	201	1	FX11_MOUSE
3	38	95.0	361	1	IHA_TRIUV
4	34	85.0	369	1	PSPD_BOVIN
5	34	85.0	371	1	CL46_BOVIN
6	34	85.0	371	1	CONG_BOVIN
7	34	85.0	6885	1	SNE1_HUMAN
8	34	85.0	9797	1	SNE1_HUMAN
9	33	82.5	201	1	FX11_HUMAN
10	33	82.5	294	1	TOXR_VIBCH
11	33	82.5	435	1	YAO6_HUMAN
12	33	82.5	500	1	AP04_HUMAN
13	33	82.5	553	1	AP01_MOUSE
14	32	80.0	247	1	MCT1_PAPHA
15	32	80.0	256	1	LFSP_PEPAM
16	32	80.0	394	1	ARAJ_ECOLI
17	32	80.0	412	1	SEPL_HUMAN
18	32	80.0	492	1	MM11_MOUSE
19	32	80.0	606	1	MM17_HUMAN
20	32	80.0	1306	1	ACE_HUMAN
21	32	80.0	1485	1	CPTX_XENLA
22	31	77.5	185	1	NPW_RAT
23	31	77.5	209	1	EPF2_MOUSE
24	31	77.5	213	1	EPF2_HUMAN
25	31	77.5	245	1	CG21_MOUSE
26	31	77.5	246	1	CG21_MOUSE
27	31	77.5	321	1	CL43_BOVIN
28	31	77.5	327	1	PPA5_RAT
29	31	77.5	328	1	IPB2_HUMAN
30	31	77.5	428	1	VRD2_ECOLI
31	31	77.5	461	1	NEB1_HUMAN
32	31	77.5	480	1	HEA1_HUMAN
33	31	77.5	530	1	PVR2_MOUSE

#### ALIGNMENTS

##### RESULT 1

ID	KLK7_HUMAN	STANDARD;	PRT;	253 AA.
AC	P49862;			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Kallikrein 7 precursor (EC 3.4.21.-) (Stratum corneum chymotryptic enzyme) (hSCCE).			
DE	KLK7 OR PRSS6 OR SCCE.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 23-53.			
RC	TISSUE=Skin;			
RC	MEDLINE=94308225; PubMed=8034709;			
RA	Hansson L., Stroemqvist M., Baeckman A., Wallbrandt P., Carlstein A.,			
RA	Egelrud T.;			
RT	"Cloning, expression, and characterization of stratum corneum			
RT	chymotryptic enzyme. A skin-specific human serine proteinase.";			
RL	J. Biol. Chem. 269:19420-19426(1994).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Keratinocytes;			
RC	Yousef G.M., Scorilas A., Diamandis E.P.;			
RT	"Molecular characterization, mapping and tissue expression of the			
RT	human stratum corneum chymotryptic enzyme gene.";			
RL	Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	MEDLINE=20510030; PubMed=11054574;			
RA	Gan L., Lee I., Smith R., Argonza-Barrett R., Lei H., McCuaig J.,			
RA	Moss P., Paepker B., Wang K.;			
RT	"Sequencing and expression analysis of the serine protease gene			
RT	cluster located in chromosome 19q13 region.";			
RL	Gene 257:119-130(2000).			
RN	[4]			
RP	SEQUENCE FROM N.A.			
RC	Hansson L., Backman A., Ny A., Edlund M., Edholm E., Tornell J.,			
RA	Wallbrandt P., Egelrud T.;			
RT	"Epidermal overexpression of stratum corneum chymotryptic enzyme in			
RT	mice: a model for chronic itchy dermatitis.";			
RL	Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.			
RN	[5]			
RP	CHARACTERIZATION.			
RC	MEDLINE=95314630; PubMed=7794273;			
RA	Skytt A., Stroemqvist M., Egelrud T.;			
RT	"Primary substrate specificity of recombinant human stratum corneum			
RT	chymotryptic enzyme.";			
RL	Biochem. Biophys. Res. Commun. 211:586-589(1995).			
CC	!- FUNCTION: May catalyze the degradation of intercellular cohesive			
CC	structures in the cornified layer of the skin in the continuous			
CC	shedding of cells from the skin surface. Specific for amino acid			
CC	residues with aromatic side chains in the P1 position. SCCE			

P51693 homo sapien  
P58401 homo sapien  
P23469 homo sapien  
Q82b56 yersinia pe  
Q9y514 homo sapien  
P08648 homo sapien  
Q9r0m0 mus musculus  
Q961q0 homo sapien  
P10273 feline leuk  
Q8sc14 pan troglod  
P31783 bos taurus  
P36565 escherichia

cleaves insulin B chain at 6-Leu-|-Cys-7, 16-Tyr-|-Leu-17, 25-Phe-|-Tyr-26 and 26-Tyr-|-Thr-27. Could play a role in the activation of precursors to inflammatory cytokines.

-!- TISSUE SPECIFICITY: Abundantly expressed in the skin and is also seen in the brain and kidney.

-!- SIMILARITY: Belongs to peptidase family S1. Kallikrein subfamily.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; L33404; AAC37551.1; -  
 EMBL; AF166330; AAD49718.1; -  
 EMBL; AF243527; AAG33360.1; -  
 EMBL; AF322593; AAK9624.1; -  
 F1R; A53968; A53968.  
 HSPF; F00763; IDPO.  
 MEROPS; S01.300; -  
 Genew; HGNC:6368; KLK7.  
 MIM; 604438; -  
 GO; GO:0008236; P:serine-type peptidase activity; TAS.  
 GO; GO:0008544; P:epidermal differentiation; TAS.  
 InterPro; IPR009003; Cys Ser trypsin.  
 InterPro; IPR001254; Peptidase\_S1.  
 InterPro; IPR001314; Peptidase\_S1A.  
 Pfam; PF00089; trypsin; 1.  
 PRINTS; PR00722; TRYPSIN.  
 SMART; SM00020; TRYP\_SPC; 1.  
 PROSITE; PS00240; TRYP\_SIN\_DOM; 1.  
 PROSITE; PS00134; TRYP\_SIN\_HIS; 1.  
 PROSITE; PS00135; TRYP\_SIN\_SER; 1.  
 W Hydrolase, Serine protease; Zymogen; Glycoprotein; Signal.  
 T SIGNAL 1 22  
 T PROPEP 23 29  
 T CHAIN 30 253  
 T ACT SITE 70 70  
 T ACT SITE 112 112  
 T ACT SITE 205 205  
 T DISULFID 36 137  
 T DISULFID 55 71  
 T DISULFID ? 239  
 T DISULFID 144 211  
 T DISULFID 176 190  
 T DISULFID 201 226  
 T CARBOHYD 246 246  
 T SEQUENCE 253 AA; 27525 MW; 2D6B6B15A76A668 CRC64;

Query Match 100.0%; Score 40; DB 1; Length 253;  
 Best Local Similarity 100.0%; Pred. No. 1;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2y 1 LLPLQILL 9  
 3b 6 LLPLQILL 14

RESULT 2  
 FK11 MOUSE  
 ID FK11 MOUSE STANDARD; PRT; 201 AA.  
 AC Q9D1M7; Q9CRB4;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE PK506 binding protein 11 precursor (EC 5.2.1.8) (Peptidyl-prolyl cis-trans isomerase) (PIase) (Rotamase) (19 kDa PK506-binding protein) (FKBP-19).  
 DE (FKBP-19).  
 DE PKBP11.  
 DE Mus musculus (Mouse).  
 OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OK NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Embryo;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,  
 RA Saito T., Okazaki Y., Gotohori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
 RA Schirai L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bult C., Pletcher C., Fujita M., Gariboldi M.,  
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-P.,  
 RA Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
 RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,  
 RA Hayashizaki Y.  
 RT "Functional annotation of a full-length mouse cDNA collection.";  
 RL Nature 409:685-690(2001).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Mammary gland;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Donald M.P., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Udwin I.B., Ioshizaki S., Carninci P., Prange C.,  
 RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Miliash S.J.,  
 RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Rulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 CC -!- FUNCTION: PIases accelerate the folding of proteins during protein synthesis.  
 CC -!- CATALYTIC ACTIVITY: Peptidylproline (omega=180) = peptidylproline (omega=0).  
 CC -!- SIMILARITY: Belongs to the FKBP-type PIase family.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; AK003331; BAB22719.1; -  
 EMBL; AK019132; BAB31559.1; -  
 EMBL; BC037596; AAK37596.1; -  
 HSPF; P20071; 1FKJ.  
 MGD; MGI:1913370; Fkbp11.  
 InterPro; IPR001179; FKBP\_PIase.  
 Pfam; PF00254; FKBP; 1.  
 PROSITE; PS00453; FKBP\_PPIASE\_1; 1.

```

PROSITE; PS00454; FKBP_PP1ASE 2; FALSE_NEG.
PROSITE; PS00559; FKBP_PP1ASE 3; 1.
ISOMERASE; Rotamase; Signal.
CHAIN 1 27 POTENTIAL.
CHAIN 28 201 FK506 BINDING PROTEIN 11.
DOMAIN 57 144 FKBP-TYPE.
CONFLICT 53 53 S > F (IN REF. 1; BAB31559).
CONFLICT 198 198 S > R (IN REF. 1; BAB31559).
SEQUENCE 201 AA; 22137 MW; 94D955C57264B082 CRC64;

Query Match 95.0%; Score 38; DB 1; Length 201;
Best Local Similarity 88.9%; Pred. No. 2.1;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

1 LLPLQLLLL 9
|||||
7 LLPLQLLLL 15

RESULT 3
HA TRIVU
D_IHA TRIVU STANDARD; PRT; 361 AA.
C 07/755;
T 15-JUL-1999 (Rel. 38, Created)
T 15-JUL-1999 (Rel. 38, Last sequence update)
T 16-OCT-2001 (Rel. 40, Last annotation update)
E Inhibin alpha chain precursor.
N INHA.
S Trichosurus vulpecula (Brush-tailed possum).
C Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
C Mammalia; Metatheria; Diprotodontia; Phalangeridae; Trichosurus.
X NCBI_TaxID=9337;
N [1]
P SEQUENCE FROM N.A.
X MEDLINE=99027340; PubMed=9801457;
A Vannontfort D., Fidler A.E., Heath D.A., Lawrence S.B., Tisdall D.J.,
LA Greenwood P.J., McNatty K.;
"CDNA sequence analysis, gene expression and protein localisation of
the inhibin alpha subunit of Australian brushtail possum (Trichosurus
vulpecula).";
J. Mol. Endocrinol. 21:141-152(1998).
C -!- FUNCTION: INHIBINS AND ACTIVINS INHIBIT AND ACTIVATE,
RESPECTIVELY, THE SECRETION OF FOLLITROPIN BY THE PITUITARY GLAND.
INHIBINS/ACTIVINS ARE INVOLVED IN REGULATING A NUMBER OF DIVERSE
FUNCTIONS SUCH AS HYPOTHALAMIC AND PITUITARY HORMONE SECRETION,
GONADAL HORMONE SECRETION, GERM CELL DEVELOPMENT AND NATURATION,
ERYTHROID DIFFERENTIATION, INSULIN SECRETION, NERVE CELL SURVIVAL,
EMBRYONIC AXIAL DEVELOPMENT OR BONE GROWTH, DEPENDING ON THEIR
SUBUNIT COMPOSITION. INHIBINS APPEAR TO OPPOSE THE FUNCTIONS OF
ACTIVINS.
C -!- SUBUNIT: DIMERIC, LINKED BY ONE OR MORE DISULFIDE BONDS.
INHIBIN A IS A DIMER OF ALPHA AND BETA-A.
INHIBIN B IS A DIMER OF ALPHA AND BETA-B.
C -!- SIMILARITY: Belongs to the TGF-beta family.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; AF033340; AAC63945.1;
GO; GO:0005576; Cytoskeletal; ISS.
GO; GO:00017106; F:activin inhibitor activity; ISS.
GO; GO:0005125; F:cyclokin activity; ISS.
GO; GO:0003793; F:defense/immunity protein activity; ISS.
GO; GO:0008083; F:growth factor activity; ISS.
GO; GO:0005179; F:hormone binding; ISS.
GO; GO:0005515; F:protein binding; ISS.
GO; GO:0007050; P:cell cycle arrest; ISS.
GO; GO:0001541; P:cell differentiation; ISS.
GO; GO:0008151; P:cell growth and/or maintenance; ISS.
GO; GO:0007166; P:cell surface receptor linked signal transdu. .; ISS.
GO; GO:0007267; P:cell-cell signaling; ISS.
GO; GO:0030218; P:erythrocyte differentiation; ISS.
GO; GO:0042541; P:hemoglobin biosynthesis; ISS.
GO; GO:0006917; P:induction of apoptosis; ISS.
GO; GO:0045578; P:negative regulation of B-cell differentiation; ISS.
GO; GO:0045786; P:negative regulation of cell cycle; ISS.
GO; GO:0046882; P:negative regulation of follicle-stimulating. .; ISS.
GO; GO:0045077; P:negative regulation of interferon-gamma bio. .; ISS.
GO; GO:0045650; P:negative regulation of macrophage different. .; ISS.
GO; GO:0042326; P:negative regulation of phosphorylation; ISS.
GO; GO:0007399; P:neurogenesis; ISS.
GO; GO:0001541; P:ovarian follicle development; ISS.
GO; GO:0046881; P:positive regulation of follicle-stimulating. .; ISS.
InterPro; IPR002405; Inhibin_alpha.
InterPro; IPR001839; TGFb.
Pfam; PF00019; TGF-beta; 1.
PRINTS; PR00669; INHIBINA.
PRODOM; PD000357; TGFb; 1.
SMART; SM00204; TGFb; 1.
PROSITE; PS00250; TGF_BETA_1; 1.
Growth factor; Hormone; Glycoprotein; Signal.
SIGNAL 1 21 POTENTIAL.
FT PROPEP 22 230 BY SIMILARITY.
FT CHAIN 231 361 INHIBIN ALPHA CHAIN.
FT DISULFID 260 323 BY SIMILARITY.
FT DISULFID 289 358 BY SIMILARITY.
FT DISULFID 293 360 BY SIMILARITY.
FT DISULFID 322 322 INTERCHAIN (BY SIMILARITY).
FT CARBOHYD 48 48 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 144 144 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 266 266 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 361 AA; 38945 MW; D661CDF93CDAA87D CRC64;

Query Match 95.0%; Score 38; DB 1; Length 361;
Best Local Similarity 88.9%; Pred. No. 3.6;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLPLQLLLL 9
|||||
DB 5 LLPLQLLLL 13

RESULT 4
PSPD BOVIN
ID_PSPD BOVIN STANDARD; PRT; 369 AA.
AC P35246;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Pulmonary surfactant-associated protein D precursor (SP-D) (PSP-D).
GN SFTPD OR SFTPD.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
[1]
SEQUENCE FROM N.A., AND SEQUENCE OF 208-247.
TISSEU=Lung;
MEDLINE=93170856; PubMed=8436402;
Lim B.L., Lu J., Reid K.B.M.;
"Structural similarity between bovine conglutinin and bovine lung
surfactant protein D and demonstration of liver as a site of
synthesis of conglutinin";
Immunology 78:159-165(1993).
C -!- FUNCTION: Contributes to the lung's defense against inhaled
microorganisms. Binds strongly maltose residues and to a lesser
extent other alpha-glucosyl moieties. It could participate in the
extracellular reorganization or turnover of pulmonary surfactant.
C -!- SUBUNIT: Oligomeric complex of 4 set of homotrimers.
C -!- SUBCELLULAR LOCATION: Extracellular.

```



[3] SEQUENCE FROM N.A.  
[4] TISSUE=Liver;  
[5] MEDLINE=9421517; PubMed=8163202;  
[6] Liou L.S., Sastri R., Hartshorn K.L., Lee Y.M., Okarma T.B.,  
[7] Tauber A.I., Sastri K.N.;  
[8] "Bovine conglutinin (BC) mRNA expressed in liver: cloning and  
[9] characterization of the BC cDNA reveals strong homology to surfactant  
[10] protein-D";  
[11] Gene 141:277-281(1994).  
[12] [4]  
[13] SEQUENCE FROM N.A.  
[14] TISSUE=Semen;  
[15] MEDLINE=94267222; PubMed=8207234;  
[16] Liou L.S., Sastri R., Hartshorn K.L., Lee Y.M., Okarma T.B.,  
[17] Tauber A.I., Sastri K.N.;  
[18] "Bovine conglutinin gene exon structure reveals its evolutionary  
[19] relationship to surfactant protein-D";  
[20] J. Immunol. 153:173-180(1994).  
[21] [5]  
[22] SEQUENCE FROM N.A.  
[23] TISSUE=Liver;  
[24] MEDLINE=94128104; PubMed=8297370;  
[25] Kawasaki N., Itoh N., Kawasaki T.;  
[26] "Gene organization and 5'-flanking region sequence of conglutinin: a  
[27] C-type mammalian lectin containing a collagen-like domain";  
[28] Biochem. Biophys. Res. Commun. 198:597-604(1994).  
[29] [6]  
[30] SEQUENCE OF 21-371.  
[31] MEDLINE=91131556; PubMed=1993651;  
[32] Lee Y.-M., Leiby K.R., Allar J., Paris K., Lerch S., Okarma T.B.;  
[33] "Primary structure of bovine conglutinin, a member of the C-type  
[34] animal lectin family";  
[35] J. Biol. Chem. 266:2715-2723(1991).  
[36] [7]  
[37] PRELIMINARY SEQUENCE OF 21-52.  
[38] MEDLINE=87184551; PubMed=3566740;  
[39] Young N.M., Leon M.A.;  
[40] "The carbohydrate specificity of conglutinin and its homology to  
[41] proteins in the hepatic lectin family";  
[42] Biochem. Biophys. Res. Commun. 143:645-651(1987).  
[43] [8]  
[44] FUNCTION: Calcium-dependent lectin-like protein which binds to a  
[45] yeast cell wall extract and immune complexes through the  
[46] complement component (C3bi). It is capable of binding nonreducing  
[47] terminal N-acetylglucosamine, mannose, and fucose residues.  
[48] [9]  
[49] SUBUNIT: Oligomeric complex of 4 set of homotrimers.  
[50] [10]  
[51] SIMILARITY: Contains 1 collagenous domain.  
[52] [11]  
[53] SIMILARITY: Contains 1 C-type lectin family domain.  
[54] [12]  
[55] This SWISS-PROT entry is copyright. It is produced through a collaboration  
[56] between the Swiss Institute of Bioinformatics and the EMBL outstation -  
[57] the European Bioinformatics Institute. There are no restrictions on its  
[58] use by non-profit institutions as long as its content is in no way  
[59] modified and this statement is not removed. Usage by and for commercial  
[60] entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
[61] or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
[62] [13]  
[63] EMBL; D14085; BAA03170.1; -  
[64] EMBL; X71774; CAA50665.1; -  
[65] EMBL; L18871; AAA20126.1; -  
[66] EMBL; U06860; AAB60624.1; -  
[67] EMBL; U06854; AAB60624.1; JOINED.  
[68] EMBL; U06855; AAB60624.1; JOINED.  
[69] EMBL; U06856; AAB60624.1; JOINED.  
[70] EMBL; U06857; AAB60624.1; JOINED.  
[71] EMBL; U06858; AAB60624.1; JOINED.  
[72] EMBL; U06859; AAB60624.1; JOINED.  
[73] EMBL; D25302; BAA04983.2; -  
[74] EMBL; D25296; BAA04983.2; JOINED.  
[75] EMBL; D25297; BAA04983.2; JOINED.  
[76] EMBL; D25298; BAA04983.2; JOINED.  
[77] EMBL; D25299; BAA04983.2; JOINED.  
[78] EMBL; D25300; BAA04983.2; JOINED.

DR EMBL; D25301; BAA04983.2; JOINED.  
DR PIR; I45878; I45878.  
DR PIR; JN0450; JN0450.  
DR HSSP; P35247; I808.  
DR InterPro; IPR008161; Clg\_helix.  
DR InterPro; IPR008160; Collagen.  
DR InterPro; IPR001304; Lectin\_C.  
DR Pfam; PF01391; Collagen; 3.  
DR Pfam; PF00059; Lectin\_C; 1.  
DR ProDom; PD000007; Clg\_helix; 1.  
DR SMART; SM00034; CLEF; 1.  
DR PROSITE; PS00615; C-TYPE LECTIN 1; 1.  
DR PROSITE; PS00615; C-TYPE LECTIN 2; 1.  
KW Lectin; Hydroxylation; Glycoprotein; Mannose-binding; Membrane;  
KW Collagen; Repeat; Calcium; Signal.  
FT SIGNAL 1 20 CONGLUTININ.  
FT CHAIN 21 371 COLLAGEN-LIKE.  
FT DOMAIN 46 216 C-TYPE LECTIN (SHORT FORM).  
FT DOMAIN 273 371 HYDROXYLATION.  
FT MOD\_RES 63 63 HYDROXYLATION.  
FT MOD\_RES 87 87 HYDROXYLATION.  
FT MOD\_RES 99 99 HYDROXYLATION.  
FT MOD\_RES 135 135 HYDROXYLATION.  
FT MOD\_RES 141 141 HYDROXYLATION.  
FT MOD\_RES 159 159 HYDROXYLATION.  
FT MOD\_RES 162 162 HYDROXYLATION.  
FT MOD\_RES 198 198 HYDROXYLATION.  
FT SITE 201 203 CELL ATTACHMENT SITE (POTENTIAL).  
FT DISULFID 275 369 BY SIMILARITY.  
FT DISULFID 347 361 BY SIMILARITY.  
FT CARBOHYD 337 337 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CONFLICT 173 173 R -> H (IN REF. 2 AND 3).  
FT CONFLICT 210 210 K -> S (IN REF. 6).  
FT CONFLICT 218 218 V -> A (IN REF. 2).  
FT CONFLICT 272 272 E -> V (IN REF. 2).  
SQ SEQUENCE 371 AA; 37994 MW; 867BB41992544B1F CRC64;  
Query Match 85.0%; Score 34; DB 1; Length 371;  
Best Local Similarity 77.8%; Pred. No. 23;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 1 LLPQLILL 9  
DB 3 LLPLSVLL 11  
RESULT 7  
SNE2\_HUMAN STANDARD; PRT; 5885 AA.  
ID SNE2\_HUMAN STANDARD; PRT; 5885 AA.  
AC Q8W4H0; Q8W4H3; Q8W4H4; Q8W4H5; Q8W4H6;  
AC Q9NU50; Q9NU51; Q9NU52; Q9NU53; Q9NU54;  
DT 10-OCT-2003 (Rel. 42, Created)  
DT 10-OCT-2003 (Rel. 42, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Nesprin 2 (Nuclear envelope spectrin repeat protein 2) (Syn-2)  
DE (Synaptic nuclear envelope protein 2) (Nucleus and actin connecting  
DE element protein) (NUANCE protein).  
DE SYNE2 OR NUA OR KIAA1011.  
GN Homo sapiens (Human).  
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 9), FUNCTION, CHARACTERIZATION, AND  
RP INTERACTION WITH F-ACTIN.  
RX MEDLINE=22113122; PubMed=12118075;  
RA Zhen Y.-Y., Libotte T., Manck M., Noegel A.A., Korenbaum E.;  
RT "NUANCE, a giant protein connecting the nucleus and actin  
RT cytoskeleton";  
RL J. Cell Sci. 115:3207-3222(2002).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORMS 4; 5 AND 7).  
RX MEDLINE=21652858; PubMed=11792814;



Zhang Q., Skepper J.N., Yang F., Davies J.D., Hegyi L., Roberts R.G., Weissberg P.L., Ellis J.A., Shanahan C.M.;  
 "Neospirins: a novel family of spectrin-repeat-containing proteins that localize to the nuclear membrane in multiple tissues.";  
 J. Cell Sci. 114:4485-4498(2001).  
 [3]  
 SEQUENCE FROM N.A. (ISOFORM 1).  
 MEDLINE=2296983; PubMed=12408964;  
 Zhang Q., Regnauth C., Greener M.J., Shanahan C.M., Roberts R.G.;  
 "The neospirins are giant actin-binding proteins, orthologous to Drosophila melanogaster muscle protein MSP-300.";  
 Genomics 80:473-481(2002).  
 [4]  
 SEQUENCE FROM N.A. (ISOFORM 6).  
 TISSUE=Testis;  
 Poustka A., Klein M., Mewes H.-W., Gassenhuber J., Wiemann S.;  
 Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.  
 [5]  
 SEQUENCE FROM N.A.  
 PubMed=12508121;  
 Heilig R., Eckenberg R., Petit J.-L., Fonknechten N., Da Silva C., Cattolico L., Levy M., Barbe V., De Berardinis V., Ureta-Vidal A., Pelletier E., Vico V., Anthouard V., Rowen L., Madan A., Qin S., Sun H., Du H., Pepin K., Artiguanave F., Robert C., Cuaud P., Bruels T., Gaillon O., Friedlander L., Sanson G., Brottier P., Cure S., Segreus B., Antere F., Samain S., Crespeau H., Abbasi N., Alich N., Boescu D., Dickhoff R., Dors M., Dubois I., Friedman C., Gouyvenoux M., James R., Madan A., Mairey-Estrada B., Mangenot S., Martins N., Menard M., Ocas S., Ratcliffe A., Shaffer T., Trask B., Vacherie B., Bellemere C., Belsler C., Besnard-Gonnet M., Batol-Mavel D., Boutard M., Briez-Silla S., Combette S., Dufosse-Laurent V., Farron C., Lechaplais C., Louesse C., Muselet D., Magdelenat G., Pateau E., Petit E., Sirvain-Trukniewicz P., Trybou A., Vega-Czarny N., Bataille F., Bluet E., Bordelais I., Dubois M., Dumont C., Guerin T., Hafray S., Hammadi R., Muanga J., Pellouin V., Robert D., Wundler E., Gauguet G., Roy A., Sainte-Marthe L., Verdier J., Verdier-Discala C., Hillier L.W., Fulton L., McPherson J., Matsuda F., Wilson R., Scarpa C., Gyapay G., Wincker P., Saurin W., Quetier F., Watson R., Hood L., Weissbach J.;  
 "The DNA sequence and analysis of human chromosome 14.";  
 Nature 421:601-607(2003).  
 [6]  
 SEQUENCE FROM N.A. (ISOFORM 8).  
 TISSUE=Brain;  
 MEDLINE=22388257; PubMed=12477932;  
 Straube R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Udén T.B., Tothyluki S., Carninci P., Prange C., Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J., Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E., Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences.";  
 Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 [7]  
 SEQUENCE FROM N.A. (ISOFORM 3), AND SEQUENCE OF 1-956 AND 5133-6885 FROM N.A.  
 TISSUE=Spleen, and Tongue;  
 Jikuya H., Takano J., Nomura N., Kikuno R., Nagase T., Ohara O., Ninomiya K., Wagatsuma M., Kanda K., Kondo H., Yokoi T., Kodaira H., Furiya T., Takahashi M., Kikkawa E., Omura Y., Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M., Yamazaki M., Sugiyama T., Irie R.,

RA Otsuki T., Sato H., Wakanatsu A., Ishii S., Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B., Suzuki Y., Sugano S., Nagahari K., Masuh Y., Nagai K., Isogai T.;  
 "NEDO human cDNA sequencing project";  
 Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.  
 [8]  
 SEQUENCE OF 5709-6885 FROM N.A. (ISOFORM 2).  
 TISSUE=Brain;  
 MEDLINE=9246063; PubMed=10231032;  
 Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Hirose M., Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;  
 "Prediction of the coding sequences of unidentified human genes. XIII. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro.";  
 DNA Res. 6:63-70(1999).  
 [9]  
 REVISIONS.  
 TISSUE=Brain;  
 MEDLINE=22158633; PubMed=12168954;  
 Nakajima D., Okazaki N., Yamakawa H., Kikuno R., Ohara O., Nagase T.;  
 "Construction of expression-ready cDNA clones for KIAA genes: manual curation of 330 KIAA cDNA clones.";  
 DNA Res. 9:99-106(2002).  
 [10]  
 SEQUENCE OF 5754-6885 FROM N.A.  
 MEDLINE=21154917; PubMed=11230166;  
 Wiemann S., Weil B., Wellenreuther R., Gassenhuber J., Glassl S., Ansoerge W., Boecker M., Bloeker H., Bauersachs S., Blum H., Lauber J., Duesterhoeft A., Beyer A., Koehler K., Strack N., Mewes H.-W., Ottenwälder B., Obermaier B., Tampe J., Heubner D., Hamblut R., Korn B., Klein M., Poustka A.;  
 "Towards a catalog of human genes and proteins: sequencing and analysis of 500 novel complete protein coding human cDNAs.";  
 Genome Res. 11:422-435(2001).  
 -1- FUNCTION: Involved in the maintenance of nuclear organization and structural integrity. Probable anchoring protein which tethers the nucleus to the cytoskeleton. Connects nuclei to the cytoskeleton by interacting with the nuclear envelope and with F-actin in the cytoplasm.  
 -1- SUBUNIT: Interacts with F-actin via its N-terminal domain.  
 -1- SUBCELLULAR LOCATION: Type IV membrane protein (Potential). The largest part of the protein is cytoplasmic, while its C-terminal part is associated with the nuclear envelope, most probably the outer nuclear membrane. Remains associated with the nuclear envelope during its breakdown in mitotic cells.  
 -1- ALTERNATIVE PRODUCTS:  
 Event=Alternative splicing; Named isoforms=9;  
 Name=1; IsoId=Q8WXH0-1; Sequence=Displayed;  
 Name=2; IsoId=Q8WXH0-2; Sequence=VSP\_007164, VSP\_007166;  
 Note=No experimental confirmation available;  
 Name=3; IsoId=Q8WXH0-3; Sequence=VSP\_007155;  
 Note=Produced by exon skipping that results in a frameshift. No experimental confirmation available;  
 Name=4; Synonyms=Beta; IsoId=Q8WXH0-4; Sequence=VSP\_007156;  
 Name=5; Synonyms=Alpha; IsoId=Q8WXH0-5; Sequence=VSP\_007157, VSP\_007164, VSP\_007165;  
 Name=6; IsoId=Q8WXH0-6; Sequence=VSP\_007158, VSP\_007165, VSP\_007166;  
 Note=No experimental confirmation available;  
 Name=7; Synonyms=Gamma; IsoId=Q8WXH0-7; Sequence=VSP\_007154, VSP\_007163;  
 Name=8; IsoId=Q8WXH0-8; Sequence=VSP\_007161, VSP\_007162;  
 Note=No experimental confirmation available;  
 Name=9; Synonyms=NUANCE-N-33; IsoId=Q8WXH0-9; Sequence=VSP\_007159, VSP\_007160;  
 -1- TISSUE SPECIFICITY: Widely expressed, with higher level in kidney.

adult and fetal liver, stomach and placenta. Weakly expressed in skeletal muscle and brain. Isoform 5 is highly expressed in pancreas, skeletal muscle and heart.

!- DOMAIN: The Klarsicht domain mediates the nuclear envelope targeting.

!- SIMILARITY: Belongs to the nesprin family.

!- SIMILARITY: Contains 1 actin-binding domain.

!- SIMILARITY: Contains 2 calponin-homology (CH) domains.

!- SIMILARITY: Contains 1 Klarsicht domain.

!- SIMILARITY: Contains 13 leucine-rich (LRR) repeats.

!- SIMILARITY: Contains 9 spectrin repeats.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

R EMBL: AF435010; AAL33547.1; -

R EMBL: AF435011; AAL33548.1; -

R EMBL: AY061757; AAL33800.1; -

R EMBL: AY061758; AAL33801.1; -

R EMBL: AY061759; AAL33802.1; -

R EMBL: AF495911; AAN60443.1; -

R EMBL: AL117404; CAB55905.1; -

R EMBL: AL162832; -; NOT\_ANNOTATED\_CDS.

R EMBL: AL355094; -; NOT\_ANNOTATED\_CDS.

Query Match 85.0%; Score 34; DB 1; Length 6885;

Best Local Similarity 87.5%; Pred. No. 4e-02; Indels 0; Gaps 0;

Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Y 2 LPQLQLLL 9

b 6837 LPQLQLLL 6844

RESULT 8

INEL HUMAN

ID SNEI HUMAN STANDARD; PRT; 8797 AA.

AC Q96F91; Q94890; Q8NP97; Q8TCF1; Q8WW66; Q8WW77; Q8MXF6; Q96N17;

AC Q9COA7; Q9HS25; Q9NS26; Q9NS36; Q9NU50; Q9UJ06; Q9UC07; Q9ULF8;

XT 10-OCT-2003 (Rel. 42, Created)

XT 10-OCT-2003 (Rel. 42, Last sequence update)

XT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Nesprin 1 (Nuclear envelope spectrin repeat protein 1) (Synaptic

DE nuclear envelope protein 1) (Syn-1) (Myocyte nuclear envelope protein

DE 1) (Myne-1) (Enaptin).

EN SYNE1 OR MYNE1 OR KIAA0796 OR KIAA1756 OR KIAA1262.

CS Homo sapiens (Human).

CC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

CX NCBI\_TaxID=9606;

RN [1]

RN SEQUENCE FROM N.A. (ISOFORMS 2 AND 3), CHARACTERIZATION, AND

RP MUTAGENESIS OF 8758-LEU-CYS-8763.

RC TISSUE=Heart, Placenta, Skeletal muscle, Spleen, and Testis;

RX MEDLINE=21652858; PubMed=11792814;

RA Zhang Q., Skepper J.N., Yang P., Davies J.D., Hegyi L., Roberts R.G.,

RA Weissberg P.L., Ellis J.A., Shanahan C.M.;

RA "Nesprins: a novel family of spectrin-repeat-containing proteins that

RT localize to the nuclear membrane in multiple tissues.";

RL J. Cell Sci. 114:4485-4498(2001).

RN [2]

RN SEQUENCE FROM N.A. (ISOFORM 1), AND SUBCELLULAR LOCATION.

RC TISSUE=Heart, Spleen, and Testis;

RX MEDLINE=22296983; PubMed=12408964;

RA Zhang O., Ragnauth C., Greener M.J., Shanahan C.M., Roberts R.G.;

RA "The nesprins are giant actin-binding proteins, orthologous to

RT Drosophila melanogaster muscle protein MSP-300.";

RL Genomics 80:473-481(2002).

RN SEQUENCE FROM N.A. (ISOFORM 4), AND VARIANT GLY-8323.

RA Braune S., Abraham S., Padmakumar V., Tunggai B., Noegel A.A.,

RA Korenbaum E.;

RT "The longest isoform of enaptin/Syne-1, a nuclear envelope associated

RT protein, binds actin cytoskeleton via the alpha-actinin-like actin-

RT binding domain.";

RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.

RN [4]

RN SEQUENCE FROM N.A. (ISOFORMS 8 AND 9).

RA Zhang Q., Shanahan C.M.;

RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.

RN [5]

RN SEQUENCE FROM N.A.

RA Almeida J., Clark S., Griffiths C., Lloyd D., Parker A., Smith M.,

RA Tracey A., Williams S.;

RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.

RN [6]

RN SEQUENCE OF 1-856 FROM N.A.

RC TISSUE=Kidney;

RA Gough L., Fan J., Lisa G., Chu S., Winnick S., Beck K.A.;

RT "Golgi localization of syne-1.";

RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.

RN [7]

RN SEQUENCE OF 28-778 AND 2901-3476 FROM N.A.

RC TISSUE=Adrenal gland, and Teratocarcinoma;

RA Ninomiya K., Wagatsuma M., Kanda K., Kondo H., Yokoi T., Kodaira H.,

RA Furuya T., Takahashi M., Kikkawa K., Omura Y., Abe K., Kamihara K.,

RA Katsuta N., Sato K., Tanikawa M., Yamazaki M., Sugiyama T., Irie R.,

RA Otaki T., Sato H., Makamatsu A., Ishii S., Yamamoto J., Isono Y.,

RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,

RA Matsuo K., Makamura Y., Sekine M., Kikuchi H., Murakawa K.,

RA Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,

RA Takemoto M., Ota T., Suzuki Y., Sugano S., Nagahari K., Masuho Y.,

RA Nagai K., Isogai T.;

RT "NED0 human CDNA sequencing project.";

RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.

RN [8]

RN SEQUENCE OF 443-8797 FROM N.A. (ISOFORM 5).

RC TISSUE=Brain;

RA Nagase T., Kikuno R., Hattori A., Kondo Y., Okumura K., Ohara O.;

RT "Prediction of the coding sequences of unidentified human genes. XIX.

RT The complete sequences of 100 new cDNA clones from brain which code

RT for large proteins in vitro.";

RL DNA Res. 7:347-355(2000).

RN [9]

RN SEQUENCE OF 743-8797 FROM N.A. (ISOFORM 6).

RC TISSUE=Brain;

RA Ansorge W., Wirkner U., Mewes H.-W., Weil B., Wiemann S.;

RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.

RN [10]

RN SEQUENCE OF 4219-8797 FROM N.A. (ISOFORM 7).

RC TISSUE=Brain;

RA Nagase T., Ishikawa K.-I., Kikuno R., Hirose M., Nomura N.,

RA Chara O.;

RT "Prediction of the coding sequences of unidentified human genes. XV.

RT The complete sequences of 100 new cDNA clones from brain which code

RT for large proteins in vitro.";

RL DNA Res. 6:337-345(1999).

RN [11]

RN SEQUENCE OF 6922-8797 FROM N.A.

RC TISSUE=Brain;

RA MEDLINE=99087487; PubMed=9872452;

RA Tanaka A., Kotani H., Nomura N., Ohara O.;

RT "Prediction of the coding sequences of unidentified human genes. XI.

RT The complete sequences of 100 new cDNA clones from brain which code

RT for large proteins in vitro.";

RL DNA Res. 5:277-286(1998).

RN [12]

RN REVISIONS.

TISSUE=Brain;  
 MEDLINE=22156633; PubMed=12168954;  
 Nakajima D., Okazaki N., Yamakawa H., Kikuno R., Ohara O., Nagase T.;  
 "Construction of expression-ready cDNA clones for KIAA genes: manual  
 curation of 330 KIAA cDNA clones.";  
 DNA Res. 9:99-106(2002).  
 [13]  
 SEQUENCE OF 7631-8797 FROM N.A., AND CHARACTERIZATION.  
 MEDLINE=21659781; PubMed=11801724;  
 Mislow J.M.K., Kim M.S., Davis D.B., McNally E.M.;  
 "Myne-1, a spectrin repeat transmembrane protein of the myocyte inner  
 nuclear membrane, interacts with lamin A/C.";  
 J. Cell Sci. 115:61-70(2002).  
 [14]  
 SEQUENCE OF 8406-8797 FROM N.A.  
 Ma F.-R., Zhu L.-P.;  
 Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.  
 -!- FUNCTION: Involved in the maintenance of nuclear organization and  
 structural integrity. Probable anchoring protein which tethers the  
 nucleus to the cytoskeleton. Connects nuclei to the cytoskeleton  
 by interacting with the nuclear envelope and with F-actin in the  
 cytoplasm.  
 -!- SUBUNIT: Interacts with MUSK, with F-actin via its N-terminal  
 domain, and with LMNA in vitro (By similarity).  
 -!- SUBCELLULAR LOCATION: Type IV membrane protein (potential). The  
 largest part of the protein is cytoplasmic, while its C-terminal  
 part is associated with the nuclear envelope, most probably the  
 outer nuclear membrane. In skeletal and smooth muscles, a  
 significant amount is found in the sarcomeres.  
 -!- ALTERNATIVE PRODUCTS:  
 Event-Alternative splicing; Named isoforms=9;  
 Name=1;  
 IsoId=Q8NF91-1; Sequence=Displayed;  
 Name=2; Synonyms=Beta;  
 IsoId=Q8NF91-2; Sequence=VSP\_007130;  
 Name=3; Synonyms=Alpha;  
 IsoId=Q8NF91-3; Sequence=VSP\_007132, VSP\_007144;  
 Name=4;  
 IsoId=Q8NF91-4; Sequence=VSP\_007134, VSP\_007139, VSP\_007140,  
 VSP\_007144;  
 Name=5;  
 IsoId=Q8NF91-5; Sequence=VSP\_007135, VSP\_007136;  
 Note=No experimental confirmation available;  
 Name=6;  
 IsoId=Q8NF91-6; Sequence=VSP\_007137, VSP\_007138;  
 Note=No experimental confirmation available;  
 Name=7;  
 IsoId=Q8NF91-7; Sequence=VSP\_007141, VSP\_007142;  
 Note=No experimental confirmation available;  
 Name=8; Synonyms=Beta 2;  
 IsoId=Q8NF91-8; Sequence=VSP\_007131;  
 Name=9; Synonyms=Alpha 2;  
 IsoId=Q8NF91-9; Sequence=VSP\_007133, VSP\_007143, VSP\_007144;  
 -!- TISSUE SPECIFICITY: Widely expressed. Highly expressed in skeletal  
 and smooth muscles, heart, spleen, and peripheral blood  
 leukocytes.  
 -!- DOMAIN: The Klarsicht domain, which contains a transmembrane  
 domain, mediates the nuclear envelope targeting.  
 -!- SIMILARITY: Belongs to the nesprin family.  
 -!- SIMILARITY: Contains 1 actin-binding domain.  
 -!- SIMILARITY: Contains 2 calponin-homology (CH) domains.  
 -!- SIMILARITY: Contains 12 HAT repeats.  
 -!- SIMILARITY: Contains 1 Klarsicht domain.  
 -!- SIMILARITY: Contains 31 spectrin repeats.  
 -!- CAUTION: Ref.5 (CAB55865, CAB55866, CAB55867 and CAB55868)  
 sequences differ from that shown due to erroneous gene model  
 prediction.  
 -!- CAUTION: Ref.7 (BAB71097) sequence differs from that shown due to  
 a chimeric cDNA.  
 -!- CAUTION: Ref.14 sequence differs from that shown due to two  
 frameshifts in positions 8412 and 8784.  
 -----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license@isb-sib.ch) (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL; AY061755; AAL33798.1; -  
 CC EMBL; AY061756; AAL33799.1; -  
 CC EMBL; AF495910; AAM60442.1; -  
 CC EMBL; AF535142; AAN03486.1; -  
 CC EMBL; AY184203; AAO27771.1; -  
 CC EMBL; AY184206; AAO27774.1; -  
 CC EMBL; AL049548; CAB55865.1; ALT\_SEQ.  
 CC EMBL; AL049548; CAB55866.1; -  
 CC EMBL; AL078582; CAB87586.1; -  
 CC EMBL; AL136079; -; NOT ANNOTATED CDS.  
 CC EMBL; AL138832; CAC16280.1; ALT\_SEQ.  
 CC EMBL; AL138832; CAC16281.1; ALT\_SEQ.  
 CC EMBL; AL357081; -; NOT ANNOTATED CDS.  
 CC EMBL; AL450401; -; NOT ANNOTATED CDS.  
 CC EMBL; AL589963; -; NOT ANNOTATED CDS.  
 CC -----  
 CC Query Match 85.0%; Score 34; DB 1; Length 8797;  
 CC Best Local Similarity 87.5%; Pred.No.5.1e-02;  
 CC Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 CC -----  
 CC QY 2 LPLQILL 9  
 CC DB 8749 LPLQILL 8756  
 CC -----  
 CC RESULT 9  
 CC FK11 HUMAN STANDARD; PRT; 201 AA.  
 CC AC Q9NYL4;  
 CC DT 28-FEB-2003 (Rel. 41, Created)  
 CC DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 CC DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 CC DE FK506 binding protein 11 precursor (EC 5.2.1.8) (Peptidyl-prolyl cis-  
 CC DE trans isomerase) (PFase) (Rotamase) (19 kDa FK506-binding protein)  
 CC DE (FKBP-19).  
 CC GN FKBP11 OR FKBP19.  
 CC OS Homo sapiens (Human).  
 CC OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 CC OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 CC OX NCBI\_TaxID=9606;  
 CC RN [1]  
 CC RP SEQUENCE FROM N.A.  
 CC RA Ruiten S., Kay J.E., Robinson C.;  
 CC RT "Identification of novel FKBP genes.";  
 CC RN Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
 CC [2]  
 CC RP SEQUENCE FROM N.A.  
 CC RA Tissue=Pancreas, and Spleen;  
 CC RX MEDLINE=22388257; PubMed=12477932;  
 CC RA Klausner R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 CC RA Strausberg R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 CC RA Atschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 CC RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh P.,  
 CC RA Diatchenko L., Marusina K., Farmer A., Rubin G.M., Hong L.,  
 CC RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 CC RA Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Prange C.,  
 CC RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
 CC RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 CC RA Villalón D.K., Murley K.C., Hale S., Garcia A.M., Gay L.J., Rulyk S.W.,  
 CC RA Richardson S., Wozney K.C., Sodergren E.J., Lu X., Gibbs R.A.,  
 CC RA Fahey J., Halton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,  
 CC RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 CC RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 CC RA Butterfield Y.S.N., Krzywicki M.I., Skalska U., Smailus D.E.,  
 CC RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;

"Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences.";

Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

-!- FUNCTION: PPIases accelerate the folding of proteins during protein synthesis.

-!- CATALYTIC ACTIVITY: Peptidylproline (omega=180) = peptidylproline (omega=0).

-!- SIMILARITY: Belongs to the FKBP-type PPIase family.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; AF238079; AAF63478.1; -

EMBL; BC027973; AAH27973.1; -

HSSP; Q00688; 1PEK.

Genew; HGNC:18624; FKBP11.

InterPro; IPR001179; FKBP\_PPIase.

Pfam; PF00254; FKBP; 1.

PROSITE; PS00453; FKBP\_PPIASE\_1; 1.

PROSITE; PS00454; FKBP\_PPIASE\_2; FALSE\_NEG.

PROSITE; PS00059; FKBP\_PPIASE\_3; 1.

Isomerase; Rotamase; Signal.

SIGNAL 1 27 POTENTIAL.

CHAIN 28 201 FK506 BINDING PROTEIN 11.

DOMAIN 57 144 PPIASE FKBP-TYPE.

SEQUENCE 201 AA; 22180 MW; 586E430B9DDCOA2 CRC64;

Query Match 82.5%; Score 33; DB 1; Length 201;

Best Local Similarity 77.8%; Pred. No. 20;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

1 LLPLQILL 9

|||||

7 LLPLHLLL 15

RESULT 10

TOXR VIBCH

TOXR VIBCH STANDARD; PRT; 294 AA.

P15755; O9KTB9;

01-APR-1990 (Rel. 14, Created)

16-OCT-2001 (Rel. 40, Last sequence update)

16-OCT-2001 (Rel. 40, Last annotation update)

Cholera toxin transcriptional activator.

TOXR OR VC0984.

Vibrio cholerae.

Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;

Vibrionaceae; Vibrio.

NCBI\_TaxID=666;

[1]\_TaxID=666;

SEQUENCE FROM N.A.

STRAIN=Classical 569B / ATCC 25870 / Serotype O1;

MEDLINE=87102874; PubMed=3802195;

Miller V.L., Taylor R.K., Mekalanos J.J.;

"Cholera toxin transcriptional activator toxR is a transmembrane DNA binding protein."

Cell 48:271-279(1987).

[2]

SEQUENCE FROM N.A.

STRAIN=El Tor N16961 / Serotype O1;

MEDLINE=20406833; PubMed=10952301;

Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,

Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,

Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,

Emolova M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,

McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,

Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,

Praser C.M.;

"DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae.";

Nature 406:477-483(2000).

[3]

SEQUENCE OF 1-11 FROM N.A.

STRAIN=El Tor E7946;

MEDLINE=5108618; PubMed=2124707;

Parsot C., Mekalanos J.J.;

"Expression of ToxR, the transcriptional activator of the virulence factors in Vibrio cholerae, is modulated by the heat shock response.";

Proc. Natl. Acad. Sci. U.S.A. 87:9898-9902(1990).

[4]

SEQUENCE OF 263-294 FROM N.A.

MEDLINE=91098651; PubMed=1898871;

Ditira V.J., Mekalanos J.J.;

"Periplasmic interaction between two membrane regulatory proteins, ToxR and ToxS, results in signal transduction and transcriptional activation.";

Cell 64:29-37(1991).

[5]

MUTAGENESIS.

MEDLINE=93015740; PubMed=1400230;

Ottemann K.M., Dirita V.J., Mekalanos J.J.;

"ToxR proteins with substitutions in residues conserved with OmpR fail to activate transcription from the cholera toxin promoter.";

J. Bacteriol. 174:6807-6814(1992).

-!- FUNCTION: THIS TRANSCRIPTION ACTIVATOR CONTROLS CHOLERA TOXIN, PLUS COLONIZATION FACTOR AND OUTER MEMBRANE PROTEIN EXPRESSION IN VIBRIO CHOLERAEE. IT BINDS TO THE 5'-TTTGAT-3' TANDEMLY REPEATED DNA SEQUENCE IN CHOLERA TOXIN PROMOTER REGION. TOXNS INTERACTS WITH THE C-TERMINAL PERIPLASMIC DOMAIN OF TOXR, STIMULATING ITS ACTIVITY. IT ACTIVATES TRANSCRIPTION AT THE PROMOTERS FOR TCP1 AND TCPA AND THIS IS PRESUMABLY VIA TOXR.

-!- SIMILARITY: TO E. COLI CADC, AND TO THE C-TERMINI OF A GROUP OF TRANSCRIPTIONAL ACTIVATORS (OMPR-LIKE PROTEINS).

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; M21249; AAA27549.1; -

EMBL; AE004179; AAP94145.1; -

EMBL; M58033; AAA27575.1; -

EMBL; M62761; AAA63558.1; -

PIR; A25970; A25970.

PIR; E82257; E82257.

TIGR; VC0984; -

InterPro; IPR001867; Trans\_reg\_C.

Pfam; PF00486; trans\_reg\_C; 1.

ProDom; PD000329; Trans\_reg\_C; 1.

Transcription regulation; DNA-binding; Activator; Transmembrane;

Trans-acting factor; Complete proteome.

DOMAIN 1 182 CYTOPLASMIC (POTENTIAL).

POTENTIAL.

TRANSMEM 183 198 PERIPLASMIC (POTENTIAL).

DOMAIN 199 294 S -> N (IN REF. 1).

CONFLICT 134 134 T -> A (IN REF. 1).

CONFLICT 172 172 T -> F (IN REF. 1).

CONFLICT 184 184 L -> F (IN REF. 1).

CONFLICT 204 204 T -> S (IN REF. 1).

SEQUENCE 294 AA; 32506 MW; 15EF85EA653C373E CRC64;

Query Match 82.5%; Score 33; DB 1; Length 294;

Best Local Similarity 77.8%; Pred. No. 29;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

1 LLPLQILL 9

|||||

190 LLPLAVLL 198

CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 DR EMBL; AL022101; CAA17877.1; -;  
 DR EMBL; AL049681; CAB41253.1; -;  
 KW Hypothetical protein.  
 FT NON TER 1  
 FT NON TER 500  
 SQ SEQUENCE 500 AA; 58247 MW; 1895CD8A8F14B7C3 CRC64;

Query Match 82.5%; Score 33; DB 1; Length 500;  
 Best Local Similarity 87.5%; Pred. No. 49;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LLPLOILL 8  
 Db 363 LVPLQILL 370

RESULT 13  
 ID APPI MOUSE STANDARD; PRT; 553 AA.  
 AC Q03157; Q8VC38;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-OCT-1993 (Rel. 27, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Amyloid-like protein 1 precursor (APLP) (APLP-1) [Contains: C30].  
 GN APLP1.  
 OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 OC NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=93066322; PubMed=1279693;  
 RA Masco W., Bupp K., Magendanz M., Gusella J.F., Tanzi R.E.,  
 RA Solomon P.;  
 RA "Identification of a mouse brain cDNA that encodes a protein related  
 RT to the Alzheimer disease-associated amyloid beta protein precursor."  
 RL Proc. Natl. Acad. Sci. U.S.A. 89:10758-10762(1992).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Retina;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan R., Moore I., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Murry D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J.J., Hilton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywicki M.I., Skalska U., Smalusz D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 human and mouse cDNA sequences."  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [3]  
 RP COLLAGEN-BINDING.  
 RX MEDLINE=96139497; PubMed=8576160;  
 RA Behr D., Hesse L., Masters C.L., Multhaup G.;  
 RT "Regulation of amyloid protein precursor (APP) binding to collagen and  
 RT mapping of the binding sites on APP and collagen type I."  
 RL J. Biol. Chem. 271:1613-1620(1996).

RESULT 11  
 ID YAO4 HUMAN STANDARD; PRT; 435 AA.  
 AC O60810;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Hypothetical protein DJ845024.5 (Fragment).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OC NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Rhodes S.;  
 RA Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
 CC -!- SIMILARITY: Belongs to the WAPE family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

IR EMBL; AL022101; CAA17880.1; -;  
 IR EMBL; AL049680; CAB41252.1; -;  
 KW Hypothetical protein.  
 FT NON TER 435  
 FT NON TER 435  
 SQ SEQUENCE 435 AA; 50539 MW; 3D19443032BBB494 CRC64;

Query Match 82.5%; Score 33; DB 1; Length 435;  
 Best Local Similarity 87.5%; Pred. No. 42;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Yy 1 LLPLOILL 8  
 Yb 298 LVPLQILL 305

RESULT 12  
 ID YAO4 HUMAN STANDARD; PRT; 500 AA.  
 AC O60810;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Hypothetical protein DJ845024.2 (Fragment).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OC NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Grafham D.;  
 RA Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.  
 CC -!- SIMILARITY: Belongs to the WAPE family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way

[4] INTERACTION WITH DAB1.  
MEDLINE=9338980; PubMed=10450257;  
Homayouni R., Rice D.S., Sheldon M., Curran T.,  
"Disabled-1 binds to the cytoplasmic domain of amyloid precursor-like  
protein 1.",  
J. Neurosci. 19:7507-7515 (1999).  
[5]  
INTERACTION WITH MAPK8IP1.  
MEDLINE=21408156; PubMed=11517249;  
Matsuda S., Yasukawa T., Homma Y., Ito Y., Niikura T., Hiraki T.,  
Hirai S., Ohno S., Kita Y., Kawasumi M., Koyama K., Yamamoto T.,  
Kyriakis J.M., Nishimoto I.,  
"C-jun N-terminal kinase (JNK)-interacting protein-1b/1slet-brain-1  
scaffolds Alzheimer's amyloid precursor protein with JNK.",  
J. Neurosci. 21:6597-6607 (2001).  
[6]  
GAMMA-SECRETASE PROCESSING, INTERACTION WITH APPB1, AND MUTAGENESIS OF  
TYR-641.  
MEDLINE=22313358; PubMed=12228233;  
Scheinfeld M.H., Ghersi E., Laky K., Fowlkes B.J., D'Adamo L.,  
"Processing of beta-amyloid precursor-like protein-1 and -2 by gamma-  
secretase regulates transcription.",  
J. Biol. Chem. 277:44195-44201 (2002).  
-!- FUNCTION: May play a role in postsynaptic function. The C-terminal  
gamma-secretase processed fragment, ALD1, activates transcription  
activation through APPB1 (Fe65) binding. Couples to JIP signal  
transduction through C-terminal binding. May interact with  
cellular G-protein signaling pathways. Can regulate neurite  
outgrowth through binding to components of the extracellular  
matrix such as heparin and collagen I.  
-!- FUNCTION: The gamma-CTF peptide, C30, is a potent enhancer of  
neuronal apoptosis (By similarity).  
-!- SUBUNIT: Binds, via its C-terminal, to the PID domain of several  
cytoplasmic proteins, including APPB and APPA family members,  
MAPK8IP1 and Dab1 (By similarity). Binding to Dab1 inhibits its  
serine phosphorylation.  
-!- SUBCELLULAR LOCATION: Type I membrane protein. C-terminally  
processed in the Golgi complex.  
-!- DOMAIN: The NPXY sequence motif found in many tyrosine-  
phosphorylated proteins is required for the specific binding of  
the PID domain. However additional amino acids either N- or C-  
terminal to the NPXY motif are often required for complete  
interaction. The NPXY site is also involved in clathrin-mediated  
endocytosis.  
-!- PTM: Proteolytically cleaved by caspases during neuronal  
apoptosis. Cleaved, in vitro, at Asp-623 by caspase-3 (By  
similarity).  
-!- PTM: N- and O-glycosylated.  
-!- MISCELLANEOUS: Binds zinc and copper in the extracellular domain.  
Zinc-binding increases heparin binding. No Cu(II) reducing  
activity with copper-binding.  
-!- SIMILARITY: Belongs to the APP family.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; L04538; AAA37247.1; -;  
EMBL; BC021877; AAH21877.1; -;  
PIR; A46362; A46362.  
DR HSSP; P05067; 1MWP.  
DR MGD; MGI:88046; APlp1.  
DR InterPro; IPR008155; A4 APP.  
DR InterPro; IPR008154; A4 extra.  
DR Pfam; PF02177; A4 EXTRA; 1.  
DR PRINTS; PR00203; AMYLOIDA.  
DR SMART; SM00006; A4 EXTRA; 1.  
PROSITE; PS00319; A4 EXTRA; 1.

DR PROSITE; PS00320; A4 INTRA; 1.  
KW Apoptosis; Endocytosis; Cell adhesion; Coated pits; Neurone;  
KW Heparin-binding; Metal-binding; Copper; Zinc; Signal; Transmembrane;  
KW Glycoprotein. 1 37 POTENTIAL.  
FT CHAIN 38 653 AMYLOID-LIKE PROTEIN 1.  
FT CHAIN 624 653 C30 (BY SIMILARITY).  
FT CHAIN 38 583 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 584 606 POTENTIAL.  
FT DOMAIN 607 653 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 157 177 COPPER-BINDING.  
FT DOMAIN 203 210 ZINC-BINDING (BY SIMILARITY).  
FT DOMAIN 313 345 HEPARIN-BINDING (BY SIMILARITY).  
FT DOMAIN 413 444 HEPARIN-BINDING (BY SIMILARITY).  
FT DOMAIN 445 462 COLLAGEN-BINDING (BY SIMILARITY).  
FT DOMAIN 263 271 POLY-GLU.  
FT DOMAIN 535 538 POLY-LEU.  
FT DOMAIN 601 606 REQUIRED FOR COPPER(II) REDUCTION (BY  
FT SITE 166 SIMILARITY).  
FT SITE 607 618 BASOLATERAL SORTING SIGNAL (BY  
FT SITE 623 624 CLEAVAGE (BY CASPASE-3) (BY SIMILARITY).  
FT SITE 641 644 ENDOCYTOSIS SIGNAL.  
FT SITE 643 646 NPXY MOTIF.  
FT CARBOHYD 464 464 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 554 554 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT MUTAGEN 641 641 Y->G: REDUCED BINDING OF APPB1.  
FT CONFLICT 17 17 P->PP (IN REF. 2).  
SQ SEQUENCE 653 AA; 72750 MW; 56516DC3EA04B0 CRC64;  
Query Match 82.5%; Score 33; DB 1; Length 653;  
Best Local Similarity 77.8%; Pred. No. 63;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 LRLPQLLL 9  
Db 22 LRLPQLLL 30  
-----  
RESULT 14  
MCT1 PAFPA  
ID MCT1 PAFPA STANDARD; PRT; 247 AA.  
AC P52195;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Chymase precursor (EC 3.4.21.39) (Mast cell protease I).  
GN CMA1 OR CHM.  
OS Papio hamadryas (Hamadryas baboon).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
OC Cercopithecoidea; Papio.  
OX NCBI\_TaxID=9557;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Liao Y., Karnik S., Husain A.;  
RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: Major secreted protease of mast cells with suspected  
roles in vasoactive peptide generation, extracellular matrix  
degradation, and regulation of gland secretion.  
CC -!- CATALYTIC ACTIVITY: Preferential cleavage: Phe-|-Xaa >  
Tyr-|-Xaa > Leu-|-Xaa.  
CC -!- SUBCELLULAR LOCATION: Mast cell granules.  
CC -!- SIMILARITY: Belongs to peptidase family S1. Granzyme subfamily.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; L04538; AAA37247.1; -;  
EMBL; BC021877; AAH21877.1; -;  
PIR; A46362; A46362.  
DR HSSP; P05067; 1MWP.  
DR MGD; MGI:88046; APlp1.  
DR InterPro; IPR008155; A4 APP.  
DR InterPro; IPR008154; A4 extra.  
DR Pfam; PF02177; A4 EXTRA; 1.  
DR PRINTS; PR00203; AMYLOIDA.  
DR SMART; SM00006; A4 EXTRA; 1.  
PROSITE; PS00319; A4 EXTRA; 1.

```
-----
EMBL; U38521; AAA91160.1; ..
EMBL; U38463; AAA91159.1; ..
HSP; P23946; 1KLT.
MEROPS; S01.140; ..
InterPro; IPR009003; Cys Ser trypsin.
InterPro; IPR001254; Peptidase_S1.
InterPro; IPR001314; Peptidase_S1A.
Pfam; PF00089; trypsin; 1.
PRINTS; PR00722; CHYMOTRYPSIN.
SMART; SMO020; TRY SP; 1.
PROSITE; PS0240; TRYPsin DOM; 1.
PROSITE; PS0134; TRYPsin HIS; 1.
PROSITE; PS0135; TRYPsin SER; 1.
Hydrolase; Serine protease; Zymogen; Signal.
FT SIGNAL 1 19
FT PROPEP 20 21
FT CHAIN 22 247
FT ACT_SITE 66 66 CHYMASE.
FT ACT_SITE 110 110 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 203 203 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 51 67 BY SIMILARITY.
FT DISULFID 144 209 BY SIMILARITY.
FT DISULFID 175 188 BY SIMILARITY.
FT CARBOHYD 80 80 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 103 103 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT SEQUENCE 247 AA; 27339 MW; E0EC15E0FA72FD8B CRC64;

Query Match 80.0%; Score 32; DB 1; Length 247;
Best Local Similarity 77.8%; Pred. No. 38;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Y 1 LLPLQILL 9
  |||||
b 3 LLPLPLLL 11
  |||||

RESULT 15
PSB PERAM STANDARD; PRT; 256 AA.
D LPSB PERAM STANDARD; PRT; 256 AA.
C P26305;
T 01-MAY-1992 (Rel. 22, Created)
T 01-MAY-1992 (Rel. 22, Last sequence update)
T 15-MAR-2004 (Rel. 43, Last annotation update)
E Hemolymph lipopolysaccharide-binding protein precursor (LPS-binding protein) (LPS-BP).
S Periplaneta americana (American cockroach).
C Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
C Blattellidae; Orthopteroidea; Dictyoptera; Blattaria; Blattodea;
C NCBI_TaxID=6975;
N [1]
P SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
C TISSUE=Hemolymph;
C MEDLINE=91302364; PubMed=1712779;
L Jomori T., Natori S.;
U "Molecular cloning of cDNA for lipopolysaccharide-binding protein from the hemolymph of the American cockroach, Periplaneta americana. Similarity of the protein with animal lectins and its acute phase expression."
U J. Biol. Chem. 266:13318-13323(1991).
C -I- FUNCTION: Participates probably in the elimination of foreign substances invading the insect abdominal cavity, and in trapping intracellular symbionts, when they leak from the mycetomes into the hemolymph.
C -I- SUBCELLULAR LOCATION: Secreted.
C -I- TISSUE SPECIFICITY: Hemolymph.
C -I- INDUCTION: By introduction of foreign cells into the abdominal cavity of adult P.americana.
C -I- MISCELLANEOUS: Calcium is required for lipopolysaccharide binding.
C -I- SIMILARITY: Contains 1 C-type lectin family domain.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
-----
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entitles requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR EMBL; D00711; BAA00616.1; ..
DR PIR; A39873; JQ0708.
DR HSSP; P20693; 1HLJ.
InterPro; IPR001304; Lectin_C.
DR Pfam; PF00059; lectin_c; 1.
DR SMART; SM00034; CLECT; 1.
DR PROSITE; PS00615; C-TYPE LECTIN_1; 1.
DR PROSITE; PS0041; C-TYPE LECTIN_2; 1.
KW Glycoprotein; Lectin; Signal; Calcium.
FT SIGNAL 1 21
FT PROPEP 22 33
FT CHAIN 34 256
FT CARBOHYD 56 56 N-LINKED (GLCNAC. . .) (PROBABLE).
FT DOMAIN 146 256 C-TYPE LECTIN (SHORT FORM).
FT DISULFID 148 252 BY SIMILARITY.
FT DISULFID 230 244 BY SIMILARITY.
FT SEQUENCE 256 AA; 28420 MW; E8DF74F1AE639858 CRC64;

Query Match 80.0%; Score 32; DB 1; Length 256;
Best Local Similarity 66.7%; Pred. No. 40;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLPLQILL 9
  |||||
Db 7 LLPLSVLL 15
  |||||

Search completed: March 1, 2004, 17:29:58
Job time : 7 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

M protein - protein search, using sw model

un on: March 1, 2004, 17:20:41 ; Search time 30.333 Seconds  
(without alignments)  
93.615 Million cell updates/sec

title: US-09-905-083-35  
erfect score: 40  
equence: 1 LLPLQILL 9

coring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

searched: 1017041 seqs, 315518202 residues

total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: SPTREMBL 25:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_ivirus:\*
- 16: sp\_bacteriap:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	40	100.0	253	Q8NSN9	Q8NSN9 homo sapien
2	38	95.0	173	Q8VCA9	Q8VCA9 mus musculus
3	38	95.0	104	Q8VSD6	Q8VSD6 mus musculus
4	36	90.0	79	Q8VB59	Q8VB59 white spot
5	36	90.0	845	Q97H76	Q97H76 clostridium
6	35	87.5	183	Q8MZZ0	Q8MZZ0 helicoverpa
7	35	87.5	208	Q21527	Q21527 letrichionom
8	35	87.5	233	Q8IMF1	Q8IMF1 drosophila
9	35	87.5	493	Q9ZKQ5	Q9ZKQ5 helicobacte
10	35	87.5	496	Q25600	Q25600 helicobacte
11	34	85.0	182	Q8MZV6	Q8MZV6 mamestra br
12	34	85.0	182	Q8MZV7	Q8MZV7 plodia inte
13	34	85.0	182	Q8MZV6	Q8MZV6 spodoptera
14	34	85.0	182	Q8MZV6	Q8MZV6 cydia pomon
15	34	85.0	242	Q7VHT6	Q7VHT6 helicobacte
16	34	85.0	256	Q922G5	Q922G5 mus musculus

17	34	85.0	262	11	Q8BMM1	Q8BMM1 mus musculus
18	34	85.0	266	11	Q8ESS3	Q8ESS3 mus musculus
19	34	85.0	267	11	Q35426	Q35426 mus musculus
20	34	85.0	267	11	Q8R1S4	Q8R1S4 rattus norv
21	34	85.0	283	11	Q8K147	Q8K147 mus musculus
22	34	85.0	311	10	Q7XIS1	Q7XIS1 oryza sativ
23	34	85.0	369	6	Q863A1	Q863A1 bos taurus
24	34	85.0	489	4	Q8N7G8	Q8N7G8 homo sapien
25	34	85.0	538	16	Q8AS1	Q8AS1 pseudomonas
26	34	85.0	539	16	Q8HTX4	Q8HTX4 pseudomonas
27	34	85.0	540	16	Q88C16	Q88C16 pseudomonas
28	34	85.0	559	16	Q82114	Q82114 streptomyce
29	34	85.0	668	4	Q86Y29	Q86Y29 homo sapien
30	34	85.0	696	16	Q7VB38	Q7VB38 prochloroco
31	34	85.0	1060	11	Q7TTP1	Q7TTP1 mus musculus
32	34	85.0	1236	2	Q9JPA4	Q9JPA4 rhodocyclus
33	34	85.0	1523	4	Q7ZJK4	Q7ZJK4 homo sapien
34	33	82.5	106	4	Q86SR8	Q86SR8 homo sapien
35	33	82.5	146	17	Q9V2D5	Q9V2D5 pyrococcus
36	33	82.5	173	8	Q9XKN1	Q9XKN1 exaerete fr
37	33	82.5	182	5	Q8MZZ8	Q8MZZ8 cydia pomon
38	33	82.5	204	16	Q8GR5	Q8GR5 pseudomonas
39	33	82.5	259	4	Q9NPP7	Q9NPP7 homo sapien
40	33	82.5	277	17	Q9HP13	Q9HP13 halobacteri
41	33	82.5	279	2	Q9352	Q9352 serratia ma
42	33	82.5	282	16	Q7Z68	Q7Z68 mycobacteri
43	33	82.5	294	16	Q53979	Q53979 mycobacteri
44	33	82.5	327	5	Q8ITE6	Q8ITE6 choristoneu
45	33	82.5	328	5	Q8ITF2	Q8ITF2 choristoneu

ALIGNMENTS

RESULT 1

Q8NSN9 ID Q8NSN9 PRELIMINARY; PRT; 253 AA.  
AC Q8NSN9; 01-OCT-2002 (TREMBLrel. 22, Created)  
DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)  
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
DE Kallikrein 7 (chymotryptic, stratum corneum).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Skin;  
RA Strauberg R.;  
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.  
DR EMBL; BC032005; AAH32005.1; -.  
DR GO; GO:0004263; F:chymotrypsin activity; IEA.  
DR GO; GO:0008233; F:peptidase activity; IEA.  
DR GO; GO:0004295; F:trypsin activity; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR InterPro; IPR009003; Cys\_Ser\_trypsin.  
DR InterPro; IPR001254; Peptidase\_S1.  
DR InterPro; IPR001314; Peptidase\_S1A.  
DR Pfam; PF00089; trypsin; 1.  
DR PRINTS; PR00722; CHYMOTRYPSIN.  
DR SMART; SM00020; Tryp\_SPC; 1.  
DR PROSITE; PS00240; TRYPSIN\_DOM; 1.  
DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
KW Hydrolase; Protease; Serine protease.  
SQ SEQUENCE 253 AA; 27608 MW; 2D68B6A41B22A668 CRC64;

Query Match 100.0%; Score 40; DB 4; Length 253;  
Best Local Similarity 100.0%; Pred. No. 4.3;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



1 LLPLQLILL 9  
|||||  
6 LLPLQLILL 14

## RESULT 2

VCAS9 PRELIMINARY; PRT; 73 AA.  
Q8VCA9  
Q8VCA9  
01-MAR-2002 (TrEMBLrel. 20, Created)  
01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
Similar to RIKEN cDNA 1110002023 gene.  
Mus musculus (Mouse).  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
NCBI\_TaxID=10090;  
[1]  
SEQUENCE FROM N.A.  
TISSUE=Salivary gland;  
Strausberg R.;  
Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.  
EMBL; BC021345; AAL21345.1; -  
SEQUENCE 73 AA; 7819 MW; 9388F5339BF3C11 CRC64;

Query Match 95.0%; Score 38; DB 11; Length 73;

Best Local Similarity 88.9%; Pred. No. 3.5;  
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

1 LLPLQLILL 9  
|||||  
7 LLPLQLILL 15

## RESULT 3

8RS5D6 PRELIMINARY; PRT; 104 AA.  
Q8RS5D6  
Q8RS5D6  
01-JUN-2002 (TrEMBLrel. 21, Created)  
01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
01-JUN-2002 (TrEMBLrel. 24, Last annotation update)  
S Similar to RIKEN cDNA 1110002023 gene.  
N FKBP11 OR 1110002023R1K.  
S Mus musculus (Mouse).  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
NCBI\_TaxID=10090;  
[1]  
SEQUENCE FROM N.A.  
Strausberg R.;  
Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
EMBL; BC022900; AAL22900.1; -  
MGD; MGI:1913370; Fkbp11.  
GO; GO:0006457; P:protein folding; IEA.  
InterPro: IPR001179; FKBP\_PPIase.  
Pfam; PF00254; FKBP; 1.  
PROSITE; PS00453; FKBP\_PPIASE 1; 1.  
PROSITE; PS00059; FKBP\_PPIASE 3; 1.  
SEQUENCE 104 AA; 11085 MW; 0534D57467566914 CRC64;

Query Match 95.0%; Score 38; DB 11; Length 104;

Best Local Similarity 88.9%; Pred. No. 4.8;  
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

1 LLPLQLILL 9  
|||||  
7 LLPLQLILL 15

## RESULT 4

Q8VB59 PRELIMINARY; PRT; 79 AA.  
ID Q8VB59  
AC Q8VB59;

DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Wav132 (WSSV187).  
OS White spot syndrome virus (WSSV).  
OC Viruses; dsDNA viruses, no RNA stage; Nimaviridae.  
OX NCBI\_TaxID=92652;  
RN [1]  
RN SEQUENCE FROM N.A.  
RP MEDLINE=21548311; PubMed=11689662;  
RX Yang F., He J., Lin X., Li Q., Pan D., Zhang X., Xu X.;  
RA "Complete genome sequence of the shrimp white spot bacilliform virus."  
RT J. Virol. 75:11811-11820(2001).  
RN [2]  
RN SEQUENCE FROM N.A.  
RA Yang F., He J., Lin X., Li Q., Pan D., Zhang X., Xu X.;  
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RN SEQUENCE FROM N.A.  
RC STRAIN=Taiwan;  
RX MEDLINE=20517548; PubMed=11062040;  
RA Tsai M.F., Yu H.T., Tzeng H.F., Leu J.H., Chou C.M., Huang C.J.,  
Wang C.H., Lin J.Y., Kou G.H., Lo C.F.;  
RT "Identification and characterization of a shrimp white spot syndrome virus (WSSV) gene that encodes a novel chimeric polypeptide of cellular-type thymidine kinase and thymidylate kinase."  
RT Virology 277:100-110(2000).  
RN [4]  
RN SEQUENCE FROM N.A.  
RC STRAIN=Taiwan;  
RX MEDLINE=21844071; PubMed=11853398;  
RA Chen L.L., Leu J.H., Huang C.J., Chou C.M., Chen S.M., Wang C.H.,  
Lo C.F., Kou G.H.;  
RT "Identification of a nucleocapsid protein (VP35) gene of shrimp white spot syndrome virus and characterization of the motif important for targeting VP35 to the nuclei of transfected insect cells."  
RT Virology 293:44-53(2002).  
RN [5]  
RN SEQUENCE FROM N.A.  
RC STRAIN=Taiwan;  
RA Lo C.-F., Kou G.-H.;  
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF332093; AAL33136.1; -  
DR EMBL; AF440570; AAL89055.1; -  
SQ SEQUENCE 79 AA; 9287 MW; 800910C7FFA0EC7E CRC64;

Query Match 90.0%; Score 36; DB 12; Length 79;

Best Local Similarity 77.8%; Pred. No. 9.5;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLPLQLILL 9

Db 34 LLPLQLILL 42

## RESULT 5

Q97H76 PRELIMINARY; PRT; 845 AA.  
ID Q97H76  
AC Q97H76;  
DT 01-OCT-2001 (TrEMBLrel. 18, Created)  
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Cation transport P-type ATPase.  
GN CAC2137.  
OS Clostridium acetobutylicum.  
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
OC Clostridium.  
OX NCBI\_TaxID=1488;  
RN [1]  
RN SEQUENCE FROM N.A.  
RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;  
RX MEDLINE=21359325; PubMed=11466286;

1 Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,  
 2 Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,  
 3 Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,  
 4 Bennett G.N., Koonin E.V., Smith D.R.;  
 5 "Genome sequence and comparative analysis of the solvent-producing  
 6 bacterium *Clostridium acetobutylicum*,"  
 7 J. Bacteriol. 183:4823-4838 (2001).  
 8 EMBL; AE007714; AAC80095.1; -;  
 9 PIR; D97163; D97163.  
 10 GO; GO:0016020; C:membrane; IEA.  
 11 GO; GO:0005524; F:ATP binding; IEA.  
 12 GO; GO:0035662; F:ATPase activity, coupled to transmembrane m. . .; IEA.  
 13 GO; GO:0016787; F:hydrolyase activity; IEA.  
 14 GO; GO:0006812; P:cation transport; IEA.  
 15 GO; GO:0008152; P:metabolism; IEA.  
 16 GO; GO:0015992; P:proton transport; IEA.  
 17 InterPro; IPR001757; ATPase\_E1-E2.  
 18 InterPro; IPR006068; Cation\_ATPase\_C.  
 19 InterPro; IPR004014; Cation\_ATPase\_N.  
 20 InterPro; IPR008250; E1-E2\_ATPase\_Reg.  
 21 InterPro; IPR005834; Hydrolyase.  
 22 InterPro; IPR000695; H\_ATPase.  
 23 Pfam; PF00689; Cation\_ATPase\_C; 1.  
 24 Pfam; PF00690; Cation\_ATPase\_N; 1.  
 25 Pfam; PF00122; E1-E2\_ATPase; 1.  
 26 Pfam; PF00702; Hydrolyase; 1.  
 27 PRINTS; PR00119; CATATPASE.  
 28 PRINTS; PR00120; HATPASE.  
 29 TIGRFAMs; TIGR01494; ATPase\_P-type; 5.  
 30 PROSITE; PS00154; ATPase\_E1-E2; 1.  
 31 Complete proteome.  
 32 Q SEQUENCE 845 AA; 93779 MW; 401293AA1FFD757 CRC64;  
 33  
 34 Query Match 90.0%; Score 36; DS 16; Length 845;  
 35 Best Local Similarity 77.8%; Pred. No. 86;  
 36 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 37  
 38 ID 1 LLPLQILL 9  
 39 :|||:|:  
 40 670 LLPLQILL 578  
 41  
 42 RESULT 6  
 43 ID Q8MZZO PRELIMINARY; PRT; 183 AA.  
 44 AC Q8MZZO;  
 45 DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
 46 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 47 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 48 DE Acyl-CoA desaturase (Fragment).  
 49 Helicoverpa assulta (Oriental tobacco budworm).  
 50 Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;  
 51 Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Noctuoidea;  
 52 Noctuidae; Heliothinae; Helicoverpa.  
 53 NCBI\_TaxID=52344;  
 54 [1]  
 55 SEQUENCE FROM N.A.  
 56 STRAIN=PPAF;  
 57 RA Knipple D.C., Rosenfield C.L., You K.M., Jeong S.E.;  
 58 RT "Evolution of the Integral Membrane Desaturase Gene Family in Moths  
 59 and Flies";  
 60 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
 61 DR EMBL; AF42910; AAM28485.1; -;  
 62 DR GO; GO:0005783; C:endoplasmic reticulum; IEA.  
 63 DR GO; GO:0016020; C:membrane; IEA.  
 64 DR GO; GO:0005506; F:iron ion binding; IEA.  
 65 DR GO; GO:0016491; F:oxidoreductase activity; IEA.  
 66 DR GO; GO:0004769; F:stearyl-CoA 9-desaturase activity; IEA.  
 67 DR GO; GO:0006633; P:fatty acid biosynthesis; IEA.  
 68 InterPro; IPR001522; Desaturase.  
 69 InterPro; IPR005804; FA\_desat\_fam.  
 70 Pfam; PF0487; FA\_desaturase; 1.  
 71 PRINTS; PR00075; FACDSDATASE.

DR ProDom: PD002221; Desaturase; 1.  
 FT NON TER 1  
 FT NON TER 183  
 SQ SEQUENCE 183 AA; 21196 MW; 728DA94FAFCD14FA CRC64;  
 Query Match 87.5%; Score 35; DB 5; Length 183;  
 Best Local Similarity 87.5%; Pred. No. 33;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 LLPLQILL 9  
 :|||:|:  
 DB 14 LLPLQILL 21  
 RESULT 7  
 ID Q21527 PRELIMINARY; PRT; 208 AA.  
 AC Q21527;  
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE NADH dehydrogenase subunit 4 (EC 1.6.5.3) (NADH-ubiquinone  
 ND4).  
 GN ND4.  
 OS Clathromomys gapperi (Southern red-backed vole).  
 OC Mitochondrion.  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Arvicolinae;  
 OC Clethrionomys;  
 OC NCBI\_TaxID=56223;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98152303; PubMed=9491603;  
 RA Engel S.R., Hogan K.M., Taylor J.F., Davis S.K.;  
 RT "Molecular systematics and paleobiogeography of the South American  
 sigmodontine rodents";  
 RL Mol. Biol. Evol. 15:35-49 (1998).  
 CC -!- CATALYTIC ACTIVITY: NADH + UBIQUINONE = NAD(+) + UBIQUINOL.  
 DR EMBL; U83808; AAB87168.1; -;  
 DR GO; GO:0005739; C:mitochondrion; IEA.  
 DR GO; GO:0008137; F:NADH dehydrogenase (ubiquinone) activity; IEA.  
 DR GO; GO:0016491; F:oxidoreductase activity; IEA.  
 DR GO; GO:0006120; P:mitochondrial electron transport, NADH to u. . .; IEA.  
 DR InterPro; IPR001918; NADHub\_extd4.  
 DR InterPro; IPR001750; Oxidored\_d1.  
 DR InterPro; IPR002260; Oxidored\_q1.  
 DR Pfam; PF00361; oxidored\_q1; 1.  
 DR Pfam; PF01059; oxidored\_q5\_N; 1.  
 DR PRINTS; PR01437; NUOXDRDTASE4.  
 KW NAD; Oxidoreductase; Ubiquinone; Mitochondrion.  
 FT NON TER 208  
 SQ SEQUENCE 208 AA; 23967 MW; 8AF1788697AED6A2 CRC64;  
 Query Match 87.5%; Score 35; DB 8; Length 208;  
 Best Local Similarity 77.8%; Pred. No. 37;  
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 LLPLQILL 9  
 :|||:|:  
 DB 99 LLPLQILL 107  
 RESULT 8  
 ID Q8MFF1 PRELIMINARY; PRT; 233 AA.  
 AC Q8MFF1;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE CG31600-PA.  
 DR CG31600.  
 GN CG31600.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Eohydroidea; Drosophilidae; Drosophila.

NCBI\_TaxID=7227;

[1]

SEQUENCE FROM N.A.  
MEDLINE=20196006; PubMed=10731132;  
Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
Ananidis P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,  
George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
Brandon R.C., Rogers Y.H., Blazewicz R.G., Champe M., Pfeiffer B.D.,  
Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,  
Abril J.F., Agbayani A., An H.J., Andrews-Pfannkuch C., Baldwin D.,  
Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
Beeson K.Y., Bencs P.V., Serman B.P., Shandari D., Bolshakov S.,  
Borkova D., Botchan M.R., Bouck J., Brockstein P., Brothier P.,  
Burtis K.C., Buesam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I.D., Dietz S.M.,  
Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
Fowler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
Harris N.H., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
Hostin D., Houson K.A., Howland T.J., Wei M.H., Ibegwam C.,  
Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
Kimmel B.B., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,  
Palazzo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,  
Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
Wang Z.Y., Wasserman D.A., Weinstock G.M., Weissbach J.,  
Williams S.M., Woodage, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,  
Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhou S., Zhu X., Smith H.O.,  
Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
T "The genome sequence of Drosophila melanogaster."  
L Science 287:2185-2195(2000).  
[2]

[3]

SEQUENCE FROM N.A.  
Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,  
Evans C.A., Gocayne J.D., Ananidis P.G., Brandon R.C., Rogers Y.,  
Bazon J., An H., Baldwin D., Bazon J., Beeson K.Y., Buesam D.A.,  
Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,  
Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,  
Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,  
Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,  
Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,  
McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunco J.,  
Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,  
Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,  
Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,  
Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;  
RT "Sequencing of Drosophila melanogaster genome."  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

[4]

SEQUENCE FROM N.A.  
Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
Hradecky P., Huang Y., Kaminer J.S., Prochnik S.E., Smith C.D.,  
Tupy J.L., Bergman C., Benham B., Carlson J.W., Celniker S.E.,  
Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,  
Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,  
Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,  
Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;  
RT "Annotation of Drosophila melanogaster genome."  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RP SEQUENCE FROM N.A.  
RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter J.C.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP SEQUENCE FROM N.A.  
RA FlyBase.  
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AE003782; AAN11138.1; -;  
DR FlyBase; FBgn0051600; CC31600.  
DR InterPro; IPR000301; Transmem. 4.  
SQ SEQUENCE 233 AA; 26529 MW; 4A6086B9238EB329 CRC64;  
Query Match 87.5%; Score 35; DB 5; Length 233;  
Best Local Similarity 77.8%; Pred. No. 41;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
QY 1 LLPLQILL 9  
Db 77 LLPLQILL 85  
[1]  
RESULT 9  
ID Q9ZKQ5 PRELIMINARY; PRT; 493 AA.  
AC Q9ZKQ5;  
DT 01-MAY-1999 (TREMELrel. 10, Created)  
DT 01-MAY-1999 (TREMELrel. 10, Last sequence update)  
DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)  
DE Putative.  
GN JUP0880.  
OS Helicobacter pylori J99 (Campylobacter pylori J99).  
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;  
OC Helicobacteraceae; Helicobacter.  
OX NCBI\_TaxID=85963;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99120557; PubMed=9923682;  
RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,  
RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,  
RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,  
RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,  
RA Trust T.J.;  
RT "Genomic sequence comparison of two unrelated isolates of the human  
gastric pathogen Helicobacter pylori."  
RL Nature 397:176-180(1999).  
DR EMBL; AE001518; AAD06464.1; -;  
DR PIR; A71875; A71875.  
DR GO; GO:0016021; C:integral to membrane; IEA.  
DR GO; GO:0015385; F:sodium:hydrogen antiporter activity; IEA.  
DR GO; GO:0008985; P:regulation of pH; IEA.  
DR GO; GO:0008914; P:sodium ion transport; IEA.  
DR InterPro; IPR004770; Antiport\_rhac.  
DR Pfam; PF03553; Na\_H\_antiporter; 1.  
KW Complete proteome.  
SQ SEQUENCE 493 AA; 53597 MW; 7F19A079A844A962 CRC64;  
Query Match 87.5%; Score 35; DB 16; Length 493;  
Best Local Similarity 88.9%; Pred. No. 83;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 LLPLQILL 9  
Db 254 LLPLQILL 262  
[3]  
RESULT 10  
ID Q25600 PRELIMINARY; PRT; 496 AA.  
AC Q25600;  
DT 01-JAN-1998 (TREMELrel. 05, Created)  
DT 01-JAN-1998 (TREMELrel. 05, Last sequence update)  
DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)  
DE Conserved hypothetical integral membrane protein.

HP0946.  
 Helicobacter pylori (Campylobacter pylori).  
 Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;  
 Helicobacteriaceae; Helicobacter.  
 [1] NCBI\_TaxID=210;  
 SEQUENCE FROM N.A.  
 STRAIN=26695 / ATCC 700392;  
 MEDLINE=97394467; PubMed=9252185;  
 Tomb J.-F., White O., Kervatage A.R., Clayton R.A., Sutton G.G.,  
 Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,  
 Nelson K., Quackenbush J., Zhou L., Kirness E.F., Peterson S.,  
 Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,  
 McKenney K., FitzGerald L.M., Lee N., Adams M.D., Hickey E.K.,  
 Berg D.E., Gocayne J.D., Uterback T.R., Peterson J.D., Kelley J.M.,  
 Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,  
 Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,  
 Venter J.C.,  
 "The complete genome sequence of the gastric pathogen Helicobacter  
 pylori.",  
 Nature 388:539-547(1997).  
 EMBL; AE000604; AAD07993.1; -.  
 F1R; B64638; B64638.  
 TIGR; HP0946; -.  
 GO; GO:0016021; C:integral to membrane; IEA.  
 GO; GO:0015385; P:sodium:hydrogen antiporter activity; IEA.  
 GO; GO:0006885; P:regulation of pH; IEA.  
 GO; GO:0006814; P:sodium ion transport; IEA.  
 InterPro; IPR004770; Antiport\_rhac.  
 Pfam; PF03553; Na H antiporter; 1.  
 Hypothetical protein; Complete proteome.  
 SEQUENCE 496 AA; 53928 MW; 3BAASDD8F14F094B CRC64;  
 Query Match 87.5%; Score 35; DB 16; Length 496;  
 Best Local Similarity 88.9%; Pred. No. 83;  
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 1 LPLQIILL 9  
 254 LPLQIILL 262

RESULT 11  
 Q8MZV6 PRELIMINARY; PRT; 182 AA.  
 ID Q8MZV6; EMBL; AF482914; AAM28498.1; -.  
 AC Q8MZV6; GO:0005783; C:endoplasmic reticulum; IEA.  
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Acyl-CoA desaturase (Fragment).  
 OS Mamestra brassicae.  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Noctuoidea;  
 OC Noctuidae; Hadeninae; Mamestra.  
 NCBI\_TaxID=55057;  
 [1]  
 SEQUENCE FROM N.A.  
 STRAIN=LPAQ;  
 Knipple D.C., Rosenfield C.L., You K.M., Jeong S.E.;  
 "Evolution of the Integral Membrane Desaturase Gene Family in Moths  
 and Flies".  
 Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
 EMBL; AF482914; AAM28498.1; -.  
 GO; GO:0005783; C:endoplasmic reticulum; IEA.  
 GO; GO:0016020; C:membrane; IEA.  
 GO; GO:0005506; P:iron ion binding; IEA.  
 GO; GO:0016491; P:oxidoreductase activity; IEA.  
 GO; GO:0004768; P:fatty acid biosynthesis; IEA.  
 GO; GO:0006633; P:fatty acid desaturase activity; IEA.  
 InterPro; IPR001522; Desaturase.  
 Pfam; PF00487; FA desaturase; 1.  
 PRINTS; PR00075; FACDSDATASE.  
 Query Match 85.0%; Score 34; DB 5; Length 182;  
 Best Local Similarity 87.5%; Pred. No. 52;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 2 LPLQIILL 9  
 14 LPLQIILL 21

RESULT 13  
 Q8MZV6 PRELIMINARY; PRT; 182 AA.  
 ID Q8MZV6; EMBL; AF482914; AAM28498.1; -.  
 AC Q8MZV6; GO:0005783; C:endoplasmic reticulum; IEA.  
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Acyl-CoA desaturase (Fragment).  
 OS Spodoptera litura.  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Noctuoidea;  
 OC Noctuidae; Amphipyrinae; Spodoptera.  
 NCBI\_TaxID=69820;  
 [1]

DR ProDom: PD002221; Desaturase; 1.  
 FT NON\_TER 1  
 FT NON\_TER 182  
 SQ SEQUENCE 182 AA; 21228 MW; 056449B6DOCC4121 CRC64;  
 Query Match 85.0%; Score 34; DB 5; Length 182;  
 Best Local Similarity 87.5%; Pred. No. 52;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 2 LPLQIILL 9  
 14 LPLQIILL 21

RESULT 12  
 Q8MZV7 PRELIMINARY; PRT; 182 AA.  
 ID Q8MZV7; EMBL; AF482923; AAM28498.1; -.  
 AC Q8MZV7; GO:0005783; C:endoplasmic reticulum; IEA.  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Acyl-CoA desaturase (Fragment).  
 OS Plodia interpunctella (Indianmeal moth).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Pyraloidea;  
 OC Pyralidae; Phycitinae; Plodia.  
 NCBI\_TaxID=58824;  
 [1]  
 SEQUENCE FROM N.A.  
 STRAIN=ASVQA;  
 Knipple D.C., Rosenfield C.L., You K.M., Jeong S.E.;  
 "Evolution of the Integral Membrane Desaturase Gene Family in Moths  
 and Flies".  
 Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
 EMBL; AF482923; AAM28498.1; -.  
 GO; GO:0005783; C:endoplasmic reticulum; IEA.  
 GO; GO:0016020; C:membrane; IEA.  
 GO; GO:0005506; P:iron ion binding; IEA.  
 GO; GO:0016491; P:oxidoreductase activity; IEA.  
 GO; GO:0004768; P:fatty acid biosynthesis; IEA.  
 GO; GO:0006633; P:fatty acid desaturase activity; IEA.  
 InterPro; IPR001522; Desaturase.  
 Pfam; PF00487; FA desaturase; 1.  
 PRINTS; PR00075; FACDSDATASE.  
 ProDom; PD002221; Desaturase; 1.  
 FT NON\_TER 1  
 FT NON\_TER 182  
 SQ SEQUENCE 182 AA; 21265 MW; 1F2F10E1A54942CC CRC64;  
 Query Match 85.0%; Score 34; DB 5; Length 182;  
 Best Local Similarity 87.5%; Pred. No. 52;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 2 LPLQIILL 9  
 14 LPLQIILL 21

RESULT 13  
 Q8MZV6 PRELIMINARY; PRT; 182 AA.  
 ID Q8MZV6; EMBL; AF482923; AAM28498.1; -.  
 AC Q8MZV6; GO:0005783; C:endoplasmic reticulum; IEA.  
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Acyl-CoA desaturase (Fragment).  
 OS Spodoptera litura.  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Noctuoidea;  
 OC Noctuidae; Amphipyrinae; Spodoptera.  
 NCBI\_TaxID=69820;  
 [1]

Db 14 LPLQLILL 21

RESULT 15

Q7VHT6 PRELIMINARY; PRT; 242 AA.

ID Q7VHT6

AC Q7VHT6

DT 01-OCT-2003 (TREMELrel. 25, Created)

DT 01-OCT-2003 (TREMELrel. 25, Last sequence update)

DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)

DE Hypothetical protein.

GN HH0877.

OS Helicobacter hepaticus.

OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;

OC Helicobacteraceae; Helicobacter.

OX NCBI\_TaxID=32025;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 51449 / 3B1;

RX MEDLINE=22709201; PubMed=12810954;

RA Suerbaum S., Josenhans C., Sterzenbach T., Drescher B., Brandt P.,

RA Bell M., Droege M., Fartmann B., Fischer H.-P., Ge Z., Hoerster A.,

RA Holland R., Klein K., Koenig J., Macko L., Mendz G.L., Nyakatura G.,

RA Schauer D.B., Shen Z., Weber J., Frosch M., Fox J.G.,

RT "The complete genome sequence of the carcinogenic bacterium

RT Helicobacter hepaticus."

RL Proc. Natl. Acad. Sci. U.S.A. 100:7901-7906(2003).

DR EMBL; AF017146; AAP77474.1; -.

KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 242 AA; 28014 MW; 5108E98D41244093 CRC64;

Query Match 85.0%; Score 34; DB 16; Length 242;

Best Local Similarity 66.7%; Pred. No. 68;

Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LPLQLILL 9

Db 221 LMPQLILL 229

Search completed: March 1, 2004, 17:34:48

Job time : 31.3333 secs

SEQUENCE FROM N.A.

STRAIN=LPSQ;

Knipple D.C., Rosenfield C.L., You K.M., Jeong S.E.;

"Evolution of the Integral Membrane Desaturase Gene Family in Moths

and Flies."

Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.

EMBL; AF482944; AAM28519.1; -.

GO; GO:0005783; C:endoplasmic reticulum; IEA.

GO; GO:0016020; C:membrane; IEA.

GO; GO:0005506; F:iron ion binding; IEA.

GO; GO:0016431; F:oxidoreductase activity; IEA.

GO; GO:0004768; F:stearoyl-CoA 9-desaturase activity; IEA.

GO; GO:0006633; P:fatty acid biosynthesis; IEA.

InterPro; IPR001522; Desaturase.

InterPro; IPR005804; FA desat. fam.

Pfam; PF00487; FA desaturase; 1.

PRINTS; PR00075; FACDSDATRASE.

ProDom; PD002221; Desaturase; 1.

NON\_TER 1

NON\_TER 182

SEQUENCE 182 AA; 21193 MW; 6E304DED17BD521B CRC64;

Query Match 85.0%; Score 34; DB 5; Length 182;

Best Local Similarity 87.5%; Pred. No. 52;

Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Y 2 LPLQLILL 9

b 14 LPLQLILL 21

RESULT 14

Q8MZ26 PRELIMINARY; PRT; 182 AA.

D Q8MZ26

C Q8MZ26

T 01-OCT-2002 (TREMELrel. 22, Created)

T 01-OCT-2002 (TREMELrel. 22, Last sequence update)

T 01-OCT-2003 (TREMELrel. 25, Last annotation update)

E Acyl-CoA desaturase (Fragment).

N SPTQ.

S Cydia pomonella (Codling moth).

C Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

C Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;

C Tortricoidae; Tortricidae; Olethreutinae; Cydia.

N NCBI\_TaxID=82600;

[1]

SEQUENCE FROM N.A.

Knipple D.C., Rosenfield C.L., You K.M., Jeong S.E.;

"Evolution of the Integral Membrane Desaturase Gene Family in Moths

and Flies."

Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.

EMBL; AF482904; AAM33374.1; -.

GO; GO:0005783; C:endoplasmic reticulum; IEA.

GO; GO:0016020; C:membrane; IEA.

GO; GO:0005506; F:iron ion binding; IEA.

GO; GO:0016431; F:oxidoreductase activity; IEA.

GO; GO:0004768; F:stearoyl-CoA 9-desaturase activity; IEA.

GO; GO:0006633; P:fatty acid biosynthesis; IEA.

InterPro; IPR001522; Desaturase.

InterPro; IPR005804; FA desat. fam.

Pfam; PF00487; FA desaturase; 1.

PRINTS; PR00075; FACDSDATRASE.

ProDom; PD002221; Desaturase; 1.

NON\_TER 1

NON\_TER 182

SEQUENCE 182 AA; 20985 MW; CFD238A73678B47A CRC64;

Query Match 85.0%; Score 34; DB 5; Length 182;

Best Local Similarity 87.5%; Pred. No. 52;

Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Y 2 LPLQLILL 9

|||||

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

M protein - protein search, using sw model

run on: March 1, 2004, 17:16:55 ; Search time 45.5556 Seconds  
(without alignments)  
55.820 Million cell updates/sec

title: US-09-905-083-36

effect score: 40  
sequence: 1 SILLPLQL 9

scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

searched: 1586107 seqs, 282547505 residues

total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: A\_Geneseq\_29Jan04:\*

- 1: Geneseq1980s:\*
- 2: Geneseq1990s:\*
- 3: Geneseq2000s:\*
- 4: Geneseq2001s:\*
- 5: Geneseq2002s:\*
- 6: Geneseq2003as:\*
- 7: Geneseq2003bs:\*
- 8: Geneseq2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

result No.	Score	Query Match	Length	ID	Description
1	40	100.0	9	AAB08241	Human str
2	40	100.0	136	ABG23378	Novel hum
3	40	100.0	136	ADA05736	Human NOV
4	40	100.0	250	ADA05732	Human NOV
5	40	100.0	253	AAR67888	Human str
6	40	100.0	253	AAW05383	Human amy
7	40	100.0	253	ABBA4421	Human SCC
8	40	100.0	253	ABBA4406	Human SCC
9	40	100.0	253	AAU82740	Amino aci
10	40	100.0	253	ABU07440	Protein d
11	40	100.0	253	ABU07471	Protein d
12	40	100.0	253	ABR58471	Human str
13	40	100.0	253	ADB0484	Ovarian c
14	40	100.0	257	ABBA1326	Human HSC
15	36	90.0	9	AAE08238	Human str
16	34	85.0	61	AAO12472	Human pol
17	33	82.5	394	ABU17803	Protein e
18	33	82.5	804	ABP70828	Human str
19	32	80.0	9	AAE08240	Human str
20	32	80.0	9	AAE08320	Human str
21	32	80.0	71	AAE08320	Human imm
22	32	80.0	73	ABG19347	Novel hum
23	32	80.0	139	ADB65713	Human pro
24	32	80.0	267	AAW22303	Rat CRTI.
25	32	80.0	282	ADD30604	Plant yie

26	31	77.5	46	3	AAB38062	Fragment
27	31	77.5	56	4	AAW79694	Human pro
28	31	77.5	89	4	ABG13602	Novel hum
29	31	77.5	89	4	ABG12769	Novel hum
30	31	77.5	142	4	AAE63580	Human gas
31	31	77.5	154	6	ABR62385	Secreted
32	31	77.5	156	4	AAE63578	Human gas
33	31	77.5	159	4	AAE63582	Human gas
34	31	77.5	189	4	ABR71026	Drosophil
35	31	77.5	190	6	ABR41229	Human DIT
36	31	77.5	211	3	AAB38058	Fragment
37	31	77.5	246	3	ABR24037	Human PRO
38	31	77.5	246	3	AAV99438	Human PRO
39	31	77.5	246	4	AAE66187	Protein o
40	31	77.5	246	6	ABO33680	Novel hum
41	31	77.5	246	7	ABO44533	Human sec
42	31	77.5	246	7	ABO33557	Novel hum
43	31	77.5	246	7	ADCI8207	Human PRO
44	31	77.5	246	7	ADD70853	Human sec
45	31	77.5	246	7	ADD39930	Human sec

## ALIGNMENTS

## RESULT 1

AAE08241  
ID AAE08241 standard; peptide; 9 AA.

AC AAE08241;

DT 01-NOV-2001 (first entry)

DE Human stratum corneum chymotrypsin enzyme peptide #6 (residues 4-12).

KW Stratum corneum chymotrypsin enzyme; SCCE; cytostatic; vaccine; tumour;  
KW cancer; ovarian; breast; lung; colon; prostate; carcinoma; human;  
KW antisenese therapy; malignant hyperplasia.

OS Homo sapiens.

PN WO200159158-A1.

PD 16-AUG-2001.

PF 07-FEB-2001; 2001WO-US003977.

PR 11-FEB-2000; 2000US-00502600.

PI (UYAR-) UNIV ARKANSAS.

XX O'brien TJ;

XX WPI; 2001-514676/56.

XX Diagnosing cancer comprises detecting stratum corneum chymotrypsin enzyme.

XX Claim 25; Page 103; 127pp; English.

XX The invention relates to diagnosing cancer especially ovarian cancer, by screening for stratum corneum chymotrypsin enzyme (SCCE). Proteases are considered to be an integral part of tumour growth and metastasis, and therefore, markers indicative of their presence or absence are useful for the diagnosis of cancer. The method is useful for diagnosing cancer, the carcinoma and malignant hyperplasia. The SCCE oligonucleotide may be used to treat a cancer selected from ovarian, breast, lung, colon, prostate and other cancers in which SCCE is overexpressed. The present sequence is human SCCE peptide

XX Sequence 9 AA;

SQ Query Match 100.0%; Score 40; DB 4; Length 9;

Best Local Similarity 100.0%; Pred. No. 1.4e+06; Mismatches 0; Indels 0; Gaps 0; Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 SLLPLQL 9  
1 SLLPLQL 9

RESULT 2  
ABG23378 standard; protein; 136 AA.  
ABG23378;  
18-FEB-2002 (first entry)  
Novel human diagnostic protein #23369.  
Human; chromosome mapping; gene mapping; gene therapy; forensic; food supplement; medical imaging; diagnostic; genetic disorder.  
Homo sapiens.  
WO200175067-A2.  
11-OCT-2001.  
30-MAR-2001; 2001WO-US008631.  
31-MAR-2000; 2000US-00540217.  
23-AUG-2000; 2000US-00649167.  
(HYSE-) HYSEQ INC.  
Drmanac RT, Liu C, Tang YT;  
WPI: 2001-639362/73.  
N-PSDB; AAS87565.  
New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity.  
Claim 20; SEQ ID NO 53737; 103pp; English.  
The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have application in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic amino acid sequences of the invention. Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://ftp.wipo.int/pub/published_pct_sequences)

Sequence 136 AA;  
Query Match 100.0%; Score 40; DB 4; Length 136;  
Best Local Similarity 100.0%; Pred. No. 5.9;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SLLPLQL 9  
Db 10 SLLPLQL 18

RESULT 3  
ADA05736  
ID ADA05736 standard; protein; 198 AA.  
XX  
AC ADA05736;  
XX  
DT 06-NOV-2003 (first entry)  
XX  
DE Human NOVI8c protein SEQ ID NO:96.  
XX  
KW human; NOVA; antidiabetic; anorectic; antibacterial; virucide; immunomodulator; cytostatic; nootropic; neuroprotective;  
KW antiparkinsonian; antilipemic; gene therapy; human disease;  
KW metabolic disorder; diabetes; obesity; infection; cachexia; cancer;  
KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;  
KW immune disorder; haematopoietic disorder; dyslipidaemia.  
XX  
OS Homo sapiens.  
XX  
PN WO2003029424-A2.  
XX  
PD 10-APR-2003.  
XX  
PF 02-OCT-2002; 2002WO-US031373.  
XX  
PR 02-OCT-2001; 2001US-0326483P.  
PR 05-OCT-2001; 2001US-0327435P.  
PR 05-OCT-2001; 2001US-0327449P.  
PR 03-OCT-2001; 2001US-0327917P.  
PR 03-OCT-2001; 2001US-0328029P.  
PR 03-OCT-2001; 2001US-0328044P.  
PR 09-OCT-2001; 2001US-0328056P.  
PR 12-OCT-2001; 2001US-0328849P.  
PR 15-OCT-2001; 2001US-0329414P.  
PR 17-OCT-2001; 2001US-0330142P.  
PR 18-OCT-2001; 2001US-0330309P.  
PR 22-OCT-2001; 2001US-0341058P.  
PR 24-OCT-2001; 2001US-0339266P.  
PR 29-OCT-2001; 2001US-0343629P.  
PR 29-OCT-2001; 2001US-0349575P.  
PR 01-NOV-2001; 2001US-0346357P.  
PR 17-APR-2002; 2002US-0373260P.  
PR 19-APR-2002; 2002US-0373815P.  
PR 19-APR-2002; 2002US-0373817P.  
PR 19-APR-2002; 2002US-0373826P.  
PR 19-APR-2002; 2002US-0373884P.  
PR 22-APR-2002; 2002US-0374977P.  
PR 16-MAY-2002; 2002US-0381037P.  
PR 16-MAY-2002; 2002US-0381038P.  
PR 16-MAY-2002; 2002US-0381042P.  
PR 17-MAY-2002; 2002US-0381642P.  
PR 28-MAY-2002; 2002US-0383656P.  
PR 25-MAY-2002; 2002US-0383656P.  
PR 25-JUN-2002; 2002US-0391335P.  
PR 01-OCT-2002; 2002US-00262511.  
(CURA-) CURAGEN CORP.  
Smithson G, Millet I, Peyman JA, Kekuda R, Ju J, Li L, Guo X;  
Patturajan M, Spytek KA, Edinger SR, Ellerman K, Malyankar UM;  
Ort T, Gorman L, Zernusen BD, Anderson DW, Zhong M, Catterton E;  
Ji W, Miller CE, Rastelli L, Stone DJ, Pena CEA, Shenoy SG;  
Shinkets RA, Rothenberg ME, Leach MD, Agee ML, Berghs C, Dipippo VA;  
Eisen AJ, Gangolli EA, Rieger DK, Spaderna SK;  
WPI; 2003-381626/36.  
N-PSDB; ADA05735.

New NOVX polypeptides and nucleic acids, useful for diagnosing, preventing or treating NOVX-associated disorders, e.g. diabetes, obesity, cancer or dyslipidemia, and in chromosome mapping, tissue typing or pharmacogenomics.

Claim 1; Page 170; 586pp; English.

The present invention describes NOVX proteins, where X can be 1 to 55 (e.g. NOV1). Also described: (1) a composition comprising a polypeptide described above and a carrier; (2) a kit comprising, in one or more containers, the composition described above; (3) an isolated nucleic acid molecule which encodes a NOVX protein of the invention; (4) a vector comprising the nucleic acid molecule described above; (5) a cell comprising the above vector; (6) an antibody that immunospecifically binds to the polypeptide described above; (7) methods for determining the presence or amount of the above polypeptide or nucleic acid molecule in a sample; (8) methods for determining the presence of or predisposition to a disease associated with altered levels of expression of the above polypeptide or nucleic acid molecule in a first mammalian subject; (9) a method of identifying an agent that binds to the polypeptide described above; (10) a method for identifying a potential therapeutic agent for use in treating a pathology that is related to an aberrant expression or aberrant physiological interactions of the polypeptide; (11) a method of screening for a modulator of activity or of latency or predisposition to a pathology associated with the polypeptide; (12) a method for modulating the activity of the polypeptide described above; (13) methods of treating or preventing a pathology associated with the above polypeptide in a mammal; and (14) a method for producing the above polypeptide. NOVX sequences have antidiabetic, anorectic, antibacterial, virucide, immunomodulator, cytostatic, nootropic, neuroprotective, antiparkinsonian and antilipidemic activities, and can be used in gene therapy. The polypeptide is useful in manufacturing a medicament for treating a syndrome associated with a human disease. The polypeptide or the nucleic acid molecule may be used to diagnose, treat or prevent metabolic disorders such as diabetes or obesity, infections, cachexia, cancer, neurodegenerative disorders such as Alzheimer's disease or Parkinson's disease, immune disorders, hematopoietic disorders and various dyslipidemias. The nucleic acids can also be used as hybridisation probes, in chromosome mapping, tissue typing, preventive medicine and pharmacogenomics. The present sequence represents a human NOVX from the present invention.

Q Sequence 198 AA;

Query Match 100.0%; Score 40; DB 6; Length 198;  
Best Local Similarity 100.0%; Pred. No. 8.6;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 SLLPLQL 9  
b 4 SLLPLQL 12

RESULT 4  
ADA05732  
X ADA05732 standard; protein; 250 AA.

X ADA05732;

X 06-NOV-2003 (first entry)

X Human NOV18a protein SEQ ID NO:92.

X human; NOVX; antidiabetic; anorectic; antibacterial; virucide;  
X immunomodulator; cytostatic; nootropic; neuroprotective;  
X antiparkinsonian; antilipidemic; gene therapy; human disease;  
X metabolic disorder; diabetes; obesity; infection; cachexia; cancer;  
X neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;  
X immune disorder; hematopoietic disorder; dyslipidemia.

X Homo sapiens.

X

WO2003029424-A2.  
XX  
PD 10-APR-2003.  
XX  
PF 02-OCT-2002; 2002WO-US031373.  
XX  
PR 02-OCT-2001; 2001US-0326483P.  
PR 05-OCT-2001; 2001US-0327435P.  
PR 05-OCT-2001; 2001US-0327449P.  
PR 09-OCT-2001; 2001US-0327917P.  
PR 09-OCT-2001; 2001US-0328029P.  
PR 09-OCT-2001; 2001US-0328044P.  
PR 09-OCT-2001; 2001US-0328056P.  
PR 12-OCT-2001; 2001US-0328849P.  
PR 15-OCT-2001; 2001US-0329414P.  
PR 17-OCT-2001; 2001US-0330142P.  
PR 18-OCT-2001; 2001US-0330309P.  
PR 22-OCT-2001; 2001US-0341058P.  
PR 24-OCT-2001; 2001US-0339266P.  
PR 24-OCT-2001; 2001US-0343629P.  
PR 29-OCT-2001; 2001US-0349575P.  
PR 01-NOV-2001; 2001US-0346357P.  
PR 17-APR-2002; 2002US-0373260P.  
PR 19-APR-2002; 2002US-0373815P.  
PR 19-APR-2002; 2002US-0373817P.  
PR 19-APR-2002; 2002US-0373826P.  
PR 19-APR-2002; 2002US-0373884P.  
PR 22-APR-2002; 2002US-0374377P.  
PR 16-MAY-2002; 2002US-0381037P.  
PR 16-MAY-2002; 2002US-0381038P.  
PR 17-MAY-2002; 2002US-0381042P.  
PR 17-MAY-2002; 2002US-0381642P.  
PR 28-MAY-2002; 2002US-0383656P.  
PR 29-MAY-2002; 2002US-0383831P.  
PR 25-JUN-2002; 2002US-0391335P.  
PR 01-OCT-2002; 2002US-00262511.  
XX (CURA-) CURAGEN CORP.

XX Smithson G, Millet I, Peyman JA, Kekuda R, Ju J, Li L, Guo X;  
XX Patturajan M, Spytek KA, Edinger SR, Ellerman K, Malyankar UM;  
XX Ott T, Gorman L, Zerkussen BD, Anderson DW, Zhong M, Catterton E;  
XX Ji W, Miller CE, Rastelli L, Stone DJ, Pena CEA, Shenoy SG;  
XX Shimkets RA, Rothenberg ME, Leach WD, Agse ML, Berghs C, DiPippo VA;  
XX Eisen AJ, Gangolli EA, Rieger DK, Spaderna SK;  
XX WFI; 2003-381626/36.  
XX N-PSDB; ADA05731.

New NOVX polypeptides and nucleic acids, useful for diagnosing, preventing or treating NOVX-associated disorders, e.g. diabetes, obesity, cancer or dyslipidemia, and in chromosome mapping, tissue typing or pharmacogenomics.

Claim 1; Page 169-170; 586pp; English.

The present invention describes NOVX proteins, where X can be 1 to 55 (e.g. NOV1). Also described: (1) a composition comprising a polypeptide described above and a carrier; (2) a kit comprising, in one or more containers, the composition described above; (3) an isolated nucleic acid molecule which encodes a NOVX protein of the invention; (4) a vector comprising the nucleic acid molecule described above; (5) a cell comprising the above vector; (6) an antibody that immunospecifically binds to the polypeptide described above; (7) methods for determining the presence or amount of the above polypeptide or nucleic acid molecule in a sample; (8) methods for determining the presence of or predisposition to a disease associated with altered levels of expression of the above polypeptide or nucleic acid molecule in a first mammalian subject; (9) a method of identifying an agent that binds to the polypeptide described above; (10) a method for identifying a potential therapeutic agent for use in treating a pathology that is related to an aberrant expression or aberrant physiological interactions of the polypeptide; (11) a method of screening for a modulator of activity or of latency or predisposition to



a pathology associated with the polypeptide; (12) a method for modulating the activity of the polypeptide described above; (13) methods of treating or preventing a pathology associated with the above polypeptide in a mammal; and (14) a method for producing the above polypeptide. NOVX sequences have antidiabetic, anorectic, antibacterial, virucide, immunomodulator, cytostatic, nootropic, neuroprotective, antiparkinsonian and antipapillary activities, and can be used in gene therapy. The polypeptide is useful in manufacturing a medicament for treating a syndrome associated with a human disease. The polypeptide or the nucleic acid molecule may be used to diagnose, treat or prevent metabolic disorders such as diabetes or obesity, infections, cachexia, cancer, neurodegenerative disorders such as Alzheimer's disease or Parkinson's disease, immune disorders, haematopoietic disorders and various dyslipidaemias. The nucleic acids can also be used as hybridisation probes, in chromosome mapping, tissue typing, preventive medicine and pharmacogenomics. The present sequence represents a human NOVX from the present invention.

Sequence 250 AA;

Query Match 100.0%; Score 40; DB 6; Length 250;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 SLLPLQL 9  
|||||  
1 SLLPLQL 9

RESULT 5

AA05383  
ID AAW05383 standard; protein; 253 AA.

AA05383

25-MAR-2003 (revised)  
09-AUG-1995 (first entry)

Human stratum corneum chymotryptic recombinant enzyme (SCCE).

Stratum corneum chymotryptic enzyme; skin disorder; acne; psoriasis;  
callosities; keratosis pilaris; ichthyoses; eczema.

Homo sapiens.

WO9500651-A1.

05-JAN-1995.

20-JUN-1994; 94WO-IB000166.

18-JUN-1993; 93DK-00000725.

(SYMBE-) SYMBICOM AB.

Egelrud T, Hansson L;

WPI; 1995-052088/07.

N-PSDB; AAQ81203.

Nucleotide sequences encoding stratum corneum chymotryptic enzyme - and related vectors, transformed cells and polypeptides, useful for treating skin disorders, e.g. acne or psoriasis, and for identification of specific inhibitors.

Disclosure; Page 97; 137pp; English.

The enzyme encoded by this sequence is used in pharmaceutical, cosmetic and skin care products, especially to treat and prevent acne, xeroderma, or other hyperkeratotic conditions (e.g. callosities or keratosis pilaris), ichthyoses, psoriasis, eczema, etc. It is produced recombinantly following mammalian, insect, plant, or microorganism transformation with plasmid pS507. (Updated on 25-MAR-2003 to correct PN

CC field.)

XX  
SQ Sequence 253 AA;

Query Match 100.0%; Score 40; DB 2; Length 253;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SLLPLQL 9  
|||||  
Db 4 SLLPLQL 12

RESULT 6

AA05383  
ID AAW05383 standard; protein; 253 AA.

XX  
AC AAW05383;

DT 31-DEC-1996 (first entry)

DE Human amyloid precursor protein protease.

XX  
KW Amyloid precursor protein protease; Alzheimer's disease; diagnosis;  
therapy.

XX  
OS Homo sapiens.

PN WO9631122-A1.

PD 10-OCT-1996.

PF 02-APR-1996; 96WO-US004294.

PR 04-APR-1995; 95US-00416257.

XX  
PA (ELIL) LILLY & CO ELI.

XX  
PI Dixon EP, Johnstone EM, Little SP;

XX  
DR WPI; 1996-464694/46.

XX  
DR N-PSDB; AAT39783.

XX  
PT New isolated human amyloid precursor protein protease - used to develop  
prods. for the treatment or diagnosis of associated conditions, esp.  
Alzheimer's disease.

XX  
PS Claim 1; Page 44-45; 55pp; English.

XX  
CC Human amyloid precursor protein protease (AAW05383) is involved in the  
processing or clearance of amyloid precursor protein to form beta-amyloid  
peptide. Its amino acid sequence was deduced from a cDNA clone (AAT39783)  
obtd. from a human lung library. Recombinant protease can be produced in  
transformed or transfected prokaryotic (partic. E. coli) or eukaryotic  
(partic. AV-120 host cells. It is used to develop products for the design  
and testing of cpds. useful for treating or preventing conditions  
associated with beta-amyloid peptide, esp. Alzheimer's disease

XX  
SQ Sequence 253 AA;

Query Match 100.0%; Score 40; DB 2; Length 253;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SLLPLQL 9  
|||||  
Db 4 SLLPLQL 12

RESULT 7

AB84421  
ID ABB84421 standard; peptide; 253 AA.

XX

C ABB84421;  
X T T (first entry)  
X Human SCCE protein N-terminal fragment SEQ ID 48.  
X SCCE, human; stratum corneum chymotryptic enzyme; kallikrein 7;  
W serine protease; transgenic mammal; skin; skin disease; skin cancer;  
W hyperkeratosis; acanthosis; epidermal inflammation; dermal inflammation;  
X pruritus; atopic dermatitis; eczema; acne; itch; KLK7; enzyme.  
X Homo sapiens.  
S WO200262135-A2.  
N D 15-AUG-2002.  
X F 08-FEB-2002; 2002WO-IB001300.  
X R 09-FEB-2001; 2001CA-02332655.  
X R 09-FEB-2001; 2001DK-00000218.  
A (EGEL/) EGELRUD T.  
A (HANS/) HANSSON L.  
X Egelrud T, Hansson L;  
X WPI; 2002-643380/69.  
X Transgenic mammal or its embryo useful as model for human disease, has  
X heterologous nucleotide sequence coding for stratum corneum chymotryptic  
X enzyme operably linked to promoter that drives its expression in skin.  
X Example 6; Page 37; 74pp; English.  
X This invention describes a novel non-human transgenic mammal or mammalian  
X embryo having integrated within its genome, a heterologous nucleotide  
X sequence comprising at least a significant part of a nucleotide sequence  
X coding for a stratum corneum chymotryptic enzyme (SCCE) or its variant,  
X operably linked to a promoter that drives expression of heterologous scce  
X or its variant in skin. The product of the invention is useful as a model  
X for the study of disease with the aim of improving treatment, to relieve  
X or ameliorate a pathogenic condition, for development or testing of a  
X cosmetic or a pharmaceutical formulation, and for the development of a  
X diagnostic method. It can also be used as a model for a skin disease or  
X skin cancer. The invention is also useful for screening or identifying a  
X compound or composition effective for the prevention or treatment of an  
X inflammatory skin diseases selected from diseases consisting of epidermal  
X hyperkeratosis, acanthosis, epidermal inflammation, dermal inflammation,  
X pruritus, atopic dermatitis, eczema, acne and inherited skin diseases  
X with epidermal hyperkeratosis. The mammal of the invention is also useful  
X as a model for further studies of itch mechanisms and the testing of  
X potential compounds and compositions for relieve of various skin diseases  
X where itch is a component. This sequence represents the N-terminal  
X fragment of the human stratum corneum chymotryptic enzyme, SCCE  
X synonymous with human kallikrein 7 (KLK7), used in the development of the  
X transgenic mammals described in the invention

Sequence 253 AA;

Query Match 100.0%; Score 40; DB 5; Length 253;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SLLPLQL 9  
| | | | | | | | | |  
DB 4 SLLPLQL 12

RESULT 8  
ABB84406

ID ABB84406 standard; protein; 253 AA.  
XX ABB84406;  
XX 08-NOV-2002 (first entry)  
DT Human SCCE protein.  
XX SCCE, human; stratum corneum chymotryptic enzyme; kallikrein 7;  
XX serine protease; transgenic mammal; skin; skin disease; skin cancer;  
KW hyperkeratosis; acanthosis; epidermal inflammation; dermal inflammation;  
KW pruritus; atopic dermatitis; eczema; acne; itch; KLK7; enzyme.  
XX Homo sapiens.  
OS WO200262135-A2.  
XX 15-AUG-2002.  
XX 08-FEB-2002; 2002WO-IB001300.  
XX 09-FEB-2001; 2001CA-02332655.  
XX 09-FEB-2001; 2001DK-00000218.  
XX (EGEL/) EGELRUD T.  
XX (HANS/) HANSSON L.  
XX Egelrud T, Hansson L;  
XX WPI; 2002-643380/69.  
XX N-PSDB; ABQ76226.  
XX Transgenic mammal or its embryo useful as model for human disease, has  
XX heterologous nucleotide sequence coding for stratum corneum chymotryptic  
XX enzyme operably linked to promoter that drives its expression in skin.  
XX Claim 10; Page 58-59; 74pp; English.  
XX This invention describes a novel non-human transgenic mammal or mammalian  
XX embryo having integrated within its genome, a heterologous nucleotide  
XX sequence comprising at least a significant part of a nucleotide sequence  
XX coding for a stratum corneum chymotryptic enzyme (SCCE) or its variant,  
XX operably linked to a promoter that drives expression of heterologous scce  
XX or its variant in skin. The product of the invention is useful as a model  
XX for the study of disease with the aim of improving treatment, to relieve  
XX or ameliorate a pathogenic condition, for development or testing of a  
XX cosmetic or a pharmaceutical formulation, and for the development of a  
XX diagnostic method. It can also be used as a model for a skin disease or  
XX skin cancer. The invention is also useful for screening or identifying a  
XX compound or composition effective for the prevention or treatment of an  
XX abnormal or unwanted phenotype, and for screening or identifying a  
XX inflammatory skin diseases selected from diseases consisting of epidermal  
XX hyperkeratosis, acanthosis, epidermal inflammation, dermal inflammation,  
XX pruritus, atopic dermatitis, eczema, acne and inherited skin diseases  
XX with epidermal hyperkeratosis. The mammal of the invention is also useful  
XX as a model for further studies of itch mechanisms and the testing of  
XX potential compounds and compositions for relieve of various skin diseases  
XX where itch is a component. This sequence represents the human stratum  
XX corneum chymotryptic enzyme, SCCE which is a serine protease synonymous  
XX with human kallikrein 7 (KLK7) and is used in the development of the  
XX transgenic mammals described in the invention

Sequence 253 AA;

Query Match 100.0%; Score 40; DB 5; Length 253;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SLLPLQL 9  
| | | | | | | | | |  
DB 4 SLLPLQL 12



Query Match 100.0%; Score 40; DB 6; Length 253;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 SLLPLQL 9  
4 SLLPLQL 12

RESULT 11  
BU07471  
ABU07471 standard; protein; 253 AA.

ABU07471;

28-JAN-2003 (first entry)

Protein differentially regulated in prostate cancer #74.

Prostate cancer; gene expression; differential regulation;  
molecular marker; drug target; cancer detection; cancer diagnosis;  
cancer staging; cancer grading; cancer assessing; cancer monitoring.

Homo sapiens.

WO200281638-A2.

17-OCT-2002.

08-APR-2002; 2002WO-US010824.

06-APR-2001; 2001US-0281731P.

06-APR-2001; 2001US-0281732P.

(ORIG-) ORIGENE TECHNOLOGIES INC.

Sun Z, Jay G;

WPI; 2003-058520/05.

N-PSDB; ABX10375.

Novel genes which are differentially regulated in prostate cancer, useful  
for diagnosing prostate cancer in prostate tissue sample and assessing  
therapeutic or preventive intervention in prostate cancer patients.

Claim 1; Page 351; 416pp; English.

The invention describes genes (I) which are differentially regulated in  
prostate cancer. (I) is useful for diagnosing a prostate cancer in a  
sample comprising prostate tissue, which involves determining the number  
of target genes which are differentially-regulated in the sample, where  
the number is indicative of the probability that the sample comprises  
prostate cancer. (I) is useful for assessing a therapeutic or preventive  
intervention in a subject having a prostate cancer, which involves  
determining the expression levels in a sample comprising prostate tissue  
of target genes which are differentially-regulated in prostate cancer.  
Preferably, the expression levels of at least 10 genes are determined.  
(I) is also useful for identifying agents that modulate a biological  
activity of a polypeptide differentially-regulated in prostate cancer  
cells, which involves contacting a polypeptide differentially-regulated  
in prostate cancer cells with a test agent under conditions effective for  
the test agent to modulate a biological activity of the polypeptide, and  
determining whether the test agent modulates the biological activity. (I)  
is useful as molecular markers, as drug targets, and for detecting,  
diagnosing, staging, grading, assessing, monitoring, prognosticating,  
preventing or treating, determining predisposition to diseases and  
conditions especially relating to prostate cancer. (I) and its expression  
products are used in the diagnostic test to assay for presence of cancer  
e.g., in tissue sections, in biopsy sample, in total RNA, in lymph, in  
blood etc. (I) is useful for assessing cancer e.g., to determine the type  
of cancer, its stage of development, the nature of genetic defect, etc.  
The polypeptide encoded by (I) can be used as target for therapy or drug  
discovery. (I) can also be used for expressing the polypeptide and thus

CC for searching specific binding partners of the polypeptide. (I) is useful  
CC in therapeutic applications to treat prostate cancer. The identification  
CC of specific genes, and groups of genes, expressed in pathways  
CC physiologically relevant to prostate cancer permits the definition of  
CC functional and disease pathways and the delineation of targets in these  
CC pathways which are useful in diagnostic, therapeutic, and clinical  
CC applications. This is the amino acid sequence of a protein differentially  
CC regulated in prostate cancer

SQ Sequence 253 AA;

Query Match 100.0%; Score 40; DB 6; Length 253;

Best Local Similarity 100.0%; Pred. No. 11;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SLLPLQL 9

Db 4 SLLPLQL 12

RESULT 12

ABR58471  
ID ABR58471 standard; protein; 253 AA.

XX ABR58471;

DT 07-JUL-2003 (first entry)

Human stratum corneum chymotryptic enzyme - ovarian cancer clone OI:676P.

Human; cytostatic; gene therapy; vaccine; cancer; ovarian cancer.

Homo sapiens.

WO2003029468-A1.

10-APR-2003.

02-OCT-2002; 2002WO-US031467.

02-OCT-2001; 2001US-0327135P.

30-MAY-2002; 2002US-0384531P.

(CORI-) CORIXA CORP.

Algate PA, Mannion J;

WPI; 2003-372001/35.

New polynucleotide and polypeptide useful for diagnosing and/or treating  
cancer, particularly ovarian cancer, and as a vaccine.

Claim 2; Page 157-158; 169pp; English.

The invention relates to a novel isolated polynucleotide. The  
polynucleotides of the invention have cytostatic activity, and may have a  
use in gene therapy, and in a vaccine. The composition and methods are  
useful in diagnosing and/or treating cancer, particularly ovarian cancer.  
The composition may also be used as a vaccine to prevent cancer. The  
present sequence is used in the exemplification of the invention

SQ Sequence 253 AA;

Query Match 100.0%; Score 40; DB 6; Length 253;

Best Local Similarity 100.0%; Pred. No. 11;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SLLPLQL 9

Db 4 SLLPLQL 12

RESULT 13

JB80484  
ADB80484 standard; protein; 253 AA.  
ADB80484;  
04-DEC-2003 (first entry)  
Ovarian cancer-associated protein #24.  
cytostatic; gene therapy; vaccine; ovarian cancer; diagnosis;  
post-operative chemotherapy; radiation therapy; tumour prognosis;  
pre-cancerous lesion detection.  
Homo sapiens.  
WO2002102235-A2.  
27-DEC-2002.  
18-JUN-2002; 2002WO-US019297.  
18-JUN-2001; 2001US-0299234P.  
27-AUG-2001; 2001US-0315287P.  
05-SEP-2001; 2001US-0317544P.  
13-NOV-2001; 2001US-0350666P.  
12-APR-2002; 2002US-0372246P.  
(EOSB-) EOS BIOTECHNOLOGY INC.  
Mack DH, Gish KC;  
WPI; 2003-167431/16.  
N-PSDB; ADB80483.  
Detecting an ovarian cancer-associated transcript in a cell from a  
patient, comprises contacting a biological sample from the patient with a  
polynucleotide that hybridizes to an ovarian cancer gene.  
Claim 13; Page 291; 332pp; English.  
The invention relates to a method of detecting an ovarian cancer-  
associated transcript in a cell from a patient, by contacting a  
biological sample from the patient with a polynucleotide that selectively  
hybridizes to a sequence at least 80% identical to any of one of 80  
nucleic acid sequences given in the specification. The method is useful  
in diagnosing ovarian cancer and in identifying and using agents and/or  
targets that inhibit ovarian cancer. The nucleic acid molecule,  
polypeptide and the antibody may also be used in detecting ovarian  
cancers, monitoring and early detection of relapse following treatment,  
monitoring response to therapy, selecting patients for post-operative  
chemotherapy or radiation therapy, in selecting mode of therapy,  
determining tumour prognosis, early detection of pre-cancerous lesions,  
and as vaccines. This sequence corresponds to one of the proteins used  
for the detection method of the invention.

XX Human HSCBE.  
DE  
XX  
KW Human; KLK-L1; KLK-L2; KLK-L3; KLK-L4; KLK-L5; KLK-L6; HSCBE;  
KW human stratum corneum chymotryptic enzyme; kallikrein-like protein;  
KW serine protease; cytostatic; cancer; prostrate cancer.  
XX  
OS Homo sapiens.  
XX  
PN WO200053776-A2.  
XX  
PD 14-SEP-2000.  
XX  
PF 09-MAR-2000; 2000WO-CA000258.  
XX  
PR 11-MAR-1999; 99US-0124260P.  
PR 01-APR-1999; 99US-0127386P.  
PR 21-JUL-1999; 99US-0144919P.  
XX  
PA (MOUN ) MOUNT SINAI HOSPITAL.  
XX  
XX Yousef GM, Diamandis EP;  
XX WPI; 2000-587440/55.  
DR  
XX  
PT New kallikrein-like (KLK-L) proteins for diagnosing and treating KLK-L  
PT protein mediated disorders, especially cancer.  
XX  
XX Example 4; Fig 17; 184pp; English.  
XX  
XX The present sequence is human stratum corneum chymotryptic enzyme  
XX (HSCBE), a member of the kallikrein multi-gene family. Kallikreins and  
XX kallikrein-like proteins are a subgroup of the serine protease enzyme  
XX family. They catalyse the selective cleavage of specific polypeptide  
XX precursors to release peptides with potent biological activity. Nucleic  
XX acids encoding kallikrein-like proteins KLK-L1, KLK-L2, KLK-L3, KLK-L4,  
XX KLK-L5 and KLK-L6 have been isolated. The proteins are useful in the  
XX treatment, monitoring and diagnosis of cancers, especially prostate  
XX cancer. They can also be used to identify a substance that can associate  
XX with or mediate the biological activity of the proteins. Antibodies can  
XX be used to treat conditions mediated by the kallikrein-like proteins  
XX  
SQ Sequence 257 AA;  
Query Match 100.0%; Score 40; DB 3; Length 257;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 SLLPLQL 9  
Db 4 SLLPLQL 12  
RESULT 15  
AAB08238  
ID AAB08238 standard; peptide; 9 AA.  
XX  
AC AAB08238;  
XX  
XX  
DT 01-NOV-2001 (first entry)  
XX  
DE Human stratum corneum chymotrypsin enzyme peptide #3 (residues 5-13).  
XX  
KW Stratum corneum chymotrypsin enzyme; SCCE; cytostatic; vaccine; tumour;  
KW cancer; ovarian; breast; lung; colon; prostate; carcinoma; human;  
KW antisense therapy; malignant hyperplasia.  
XX  
OS Homo sapiens.  
XX  
PN WO200159158-A1.  
XX  
PD 16-AUG-2001.  
XX

07-FEB-2001; 2001WO-US003977.

11-FEB-2000; 2000US-00502600.

(UYAR-) UNIV ARKANSAS.

O'Brien TJ;

WPI; 2001-514676/56.

Diagnosing cancer comprises detecting stratum corneum chymotrypsin enzyme.

Claim 25; Page 102; 127pp; English.

The invention relates to diagnosing cancer especially ovarian cancer, by screening for stratum corneum chymotrypsin enzyme (SCCE). Proteases are considered to be an integral part of tumour growth and metastasis, and therefore, markers indicative of their presence or absence are useful for the diagnosis of cancer. The method is useful for diagnosing cancer, carcinoma and malignant hyperplasia. The SCCE oligonucleotide may be used to treat a cancer selected from ovarian, breast, lung, colon, prostate and other cancers in which SCCE is overexpressed. The present sequence is human SCCE peptide

Sequence 9 AA;

Query Match 90.0%; Score 36; DB 4; Length 9;

Best Local Similarity 100.0%; Pred.No. 1.4e+06;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 2 LLLPLQL 9

b 1 LLLPLQL 8

search completed: March 1, 2004, 17:28:53

ob time : 46.5556 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

M protein - protein search, using sw model

run on: March 1, 2004, 17:21:46 ; Search time 11.8889 Seconds  
(without alignments)  
39.081 Million cell updates/sec

title: US-09-905-083-36  
effect score: 40  
sequence: 1 SLLPLQL 9

coring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

searched: 389414 seqs, 51625971 residues

total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/prodata/2/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/prodata/2/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/prodata/2/iaa/5A\_COMB.pep:\*  
4: /cgn2\_6/prodata/2/iaa/5B\_COMB.pep:\*  
5: /cgn2\_6/prodata/2/iaa/5A\_COMB.pep:\*  
6: /cgn2\_6/prodata/2/iaa/5B\_COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	ID	Description
1	40	100.0	9	3	US-09-502-600-36
2	40	100.0	9	4	US-09-918-243-36
3	40	100.0	253	2	US-08-557-146-2
4	40	100.0	253	2	US-08-824-874-3
5	40	100.0	253	2	US-09-154-344-2
6	40	100.0	253	3	US-08-930-188-2
7	40	100.0	253	3	US-09-210-084-3
8	40	100.0	253	4	US-09-764-762-3
9	40	100.0	253	5	FCT-US96-04294-2
10	36	90.0	9	3	US-09-502-600-33
11	36	90.0	9	4	US-09-918-243-33
12	34	85.0	812	4	US-09-489-039A-12075
13	32	80.0	9	3	US-09-502-600-35
14	32	80.0	9	3	US-09-502-600-116
15	32	80.0	9	4	US-09-918-243-35
16	32	80.0	9	4	US-09-918-243-116
17	32	80.0	557	4	US-09-489-039A-10804
18	31	77.5	181	4	US-09-543-681A-5937
19	31	77.5	307	3	US-09-177-249-13
20	31	77.5	472	4	US-09-543-681A-5049
21	31	77.5	1279	4	US-09-170-496D-293
22	31	77.5	1279	4	US-03-364-425B-58
23	30	75.0	135	2	US-08-383-821-5
24	30	75.0	135	3	US-08-459-906-5
25	30	75.0	169	4	US-09-328-352-7861
26	30	75.0	231	4	US-09-252-991A-22663
27	30	75.0	308	4	US-09-252-991A-27921

28	30	75.0	404	4	US-09-489-039A-8947	Sequence 8947, Ap
29	30	75.0	436	4	US-09-489-039A-11785	Sequence 11785, A
30	30	75.0	463	2	US-08-577-049-6	Sequence 6, Appli
31	30	75.0	474	4	US-09-489-039A-10791	Sequence 10791, A
32	30	75.0	542	4	US-09-489-847-323	Sequence 323, App
33	30	75.0	549	2	US-08-576-279-57	Sequence 57, Appl
34	30	75.0	1098	1	US-07-777-715-7	Sequence 7, Appli
35	30	75.0	1098	1	US-08-170-126-2	Sequence 2, Appli
36	30	75.0	1098	3	US-08-954-418-2	Sequence 2, Appli
37	29	72.5	21	3	US-08-753-007A-24	Sequence 24, Appl
38	29	72.5	21	3	US-09-398-496-24	Sequence 22, Appl
39	29	72.5	22	4	US-08-977-378-22	Sequence 40, Appl
40	29	72.5	26	3	US-08-630-172-40	Sequence 40, Appl
41	29	72.5	26	3	US-09-375-419-40	Sequence 19, Appl
42	29	72.5	190	1	US-08-339-152A-19	Sequence 6, Appli
43	29	72.5	190	2	US-08-007-999B-6	Sequence 6, Appli
44	29	72.5	190	2	US-08-689-276A-6	Sequence 6, Appli
45	29	72.5	195	4	US-09-543-681A-6227	Sequence 6227, Ap

ALIGNMENTS

RESULT 1  
US-09-502-600-36  
; Sequence 36, Application US/09502600A  
; Patent No. 6294344  
; GENERAL INFORMATION:  
; APPLICANT: O'Brien, Timothy J.  
; TITLE OF INVENTION: Compositions and Methods for the Early Diagnosis of  
; FILE OF INVENTION: D6223CIP-C  
; FILE REFERENCE: D6223CIP-C  
; CURRENT FILING DATE: 2000-02-11  
; CURRENT APPLICATION NUMBER: US/09/502,600A  
; PRIOR APPLICATION NUMBER: 09/039,211  
; PRIOR FILING DATE: 03-14-1998  
; NUMBER OF SEQ ID NOS: 136  
; SEQ ID NO 36  
; LENGTH: 9  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: Residues 4-12 of the SCCE protein  
US-09-502-600-36

Query Match 100.0%; Score 40; DB 3; Length 9;  
Best Local Similarity 100.0%; Pred. No. 3e+05;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SLLPLQL 9  
DB 1 SLLPLQL 9

RESULT 2  
US-09-918-243-36  
; Sequence 36, Application US/09918243  
; Patent No. 6627403  
; GENERAL INFORMATION:  
; APPLICANT: O'Brien, Timothy J.  
; APPLICANT: Cannon, Martin J.  
; APPLICANT: Santin, Alessandro  
; TITLE OF INVENTION: Methods for the early diagnosis of ovarian cancer  
; FILE REFERENCE: D6223CIP/C/D/CIP  
; CURRENT APPLICATION NUMBER: US/09/918,243  
; CURRENT FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US  
; PRIOR FILING DATE: 2001-07-13  
; NUMBER OF SEQ ID NOS: 136  
; SEQ ID NO 36  
; LENGTH: 9  
; TYPE: PRT  
; ORGANISM: Homo sapiens

## FEATURE:

NAME/KEY: CHAIN  
OTHER INFORMATION: Residues 4-12 of the SCCE protein  
-09-918-243-36

Query Match 100.0%; Score 40; DB 4; Length 9;  
Best Local Similarity 100.0%; Pred. No. 3e+05;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 SLLPLQL 9  
|||||  
1 SLLPLQL 9

## RESULT 3

-08-557-146-2  
Sequence 2, Application US/08557146  
Patent No. 5834290  
GENERAL INFORMATION:  
APPLICANT: Egelrud, Torbjorn  
TITLE OF INVENTION: Recombinant Stratum Corneum Chymotryptic  
TITLE OF INVENTION: Enzyme (SCCE)  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: White & Case, Patent Department  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2787  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/557,146  
FILING DATE: 14-DEC-1995  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Sterner, Richard J.  
REGISTRATION NUMBER: 35,372  
REFERENCE/DOCKET NUMBER: 1103326-181  
TELEPHONE: (212) 819-8783  
TELEFAX: (212) 354-8113  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 253 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein

## IS-08-557-146-2

Query Match 100.0%; Score 40; DB 2; Length 253;  
Best Local Similarity 100.0%; Pred. No. 1.8;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 SLLPLQL 9  
|||||  
4 SLLPLQL 12

## RESULT 4

IS-08-824-874-3  
Sequence 3, Application US/08824874  
Patent No. 5962300  
GENERAL INFORMATION:  
APPLICANT: Hillman, Jennifer L.  
APPLICANT: Lal, Preeti  
TITLE OF INVENTION: NOVEL KALLIKREIN  
NUMBER OF SEQUENCES: 5

CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: Fast-Seq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/824,874  
FILING DATE: Filed Herewith  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0252 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 253 amino acids  
TYPE: amino acid  
STRADEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 532504  
US-08-824-874-3

Query Match 100.0%; Score 40; DB 2; Length 253;  
Best Local Similarity 100.0%; Pred. No. 1.8;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SLLPLQL 9  
|||||  
DB 4 SLLPLQL 12

## RESULT 5

US-09-154-344-2  
Sequence 2, Application US/09154344  
Patent No. 5981256  
GENERAL INFORMATION:  
APPLICANT: Egelrud, Torbjorn  
APPLICANT: Hansson, Lennart  
TITLE OF INVENTION: Recombinant Stratum Corneum Chymotryptic  
TITLE OF INVENTION: Enzyme (SCCE)  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: White & Case, Patent Department  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2787  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/154,344  
FILING DATE: 16-SEP-1998  
CLASSIFICATION:  
PRIOR APPLICATION DATA:



APPLICATION NUMBER: US 08/557,146  
FILING DATE: 14-DEC-1995  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Sterner, Richard J.  
REGISTRATION NUMBER: 35,372  
REFERENCE/DOCKET NUMBER: 1103326-181  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 819-8783  
TELEFAX: (212) 354-8113  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 253 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
3-09-154-344-2

Query Match 100.0%; Score 40; DB 2; Length 253;  
Best Local Similarity 100.0%; Pred. No. 1.8;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 SLLPLQL 9  
b 4 SLLPLQL 12

RESULT 6  
S-08-930-188-2  
Sequence 2, Application US/08930188  
Patent No. 6093397  
GENERAL INFORMATION:  
APPLICANT: Dixon, Eric P.  
APPLICANT: Johnstone, Edward M.  
APPLICANT: Little, Sheila P.  
TITLE OF INVENTION: AMYLOID PRECURSOR PROTEIN PROTEASE AND  
RELATED NUCLEIC ACIDS  
NUMBER OF SEQUENCES: 3

CORRESPONDENCE ADDRESS:  
ADDRESSEE: Eli Lilly and Company  
STREET: Lilly Corporate Center  
CITY: Indianapolis  
STATE: Indiana  
COUNTRY: United States of America  
ZIP: 46285  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/930,188  
FILING DATE:  
CLASSIFICATION:

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/416,257  
FILING DATE: 04-APR-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Blalock, Donna K.  
REGISTRATION NUMBER: 38,082  
REFERENCE/DOCKET NUMBER: X9239  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 317-277-1090  
TELEFAX: 317-276-3861  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 253 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
S-08-930-188-2

Query Match 100.0%; Score 40; DB 3; Length 253;  
Best Local Similarity 100.0%; Pred. No. 1.8;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SLLPLQL 9  
Db 4 SLLPLQL 12

RESULT 7  
US-09-210-084-3  
Sequence 3, Application US/09210084  
Patent No. 6197511  
GENERAL INFORMATION:  
APPLICANT: Hillman, Jennifer L.  
APPLICANT: Lal, Preeti  
TITLE OF INVENTION: NOVEL KALLIKREIN  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/210,084  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/824,874  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0252 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 253 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 532504

US-09-210-084-3

Query Match 100.0%; Score 40; DB 3; Length 253;  
Best Local Similarity 100.0%; Pred. No. 1.8;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SLLPLQL 9  
Db 4 SLLPLQL 12

RESULT 8  
US-09-764-762-3  
Sequence 3, Application US/09764762  
Patent No. 6472195  
GENERAL INFORMATION:  
APPLICANT: Hillman, Jennifer L.  
TITLE OF INVENTION: NOVEL KALLIKREIN  
NUMBER OF SEQUENCES: 5

CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/764,762  
FILING DATE: 16-Jan-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/210,084  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0252 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 253 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 522504  
SEQUENCE DESCRIPTION: SEQ ID NO: 3:  
S-09-764-762-3  
Query Match 100.0%; Score 40; DB 4; Length 253;  
Best Local Similarity 100.0%; Pred. No. 1.8;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Y 1 SLLPLQIL 9  
b 4 SLLPLQIL 12  
CT-US96-04294-2  
Sequence 2, Application PC/TUS9604294  
GENERAL INFORMATION:  
APPLICANT: Dixon, Eric P.  
APPLICANT: Johnstone, Edward M.  
APPLICANT: Little, Sheila P.  
TITLE OF INVENTION: AMYLOID PRECURSOR PROTEIN PROTEASE AND  
RELATED NUCLEIC ACIDS  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Eli Lilly and Company  
STREET: Lilly Corporate Center  
CITY: Indianapolis  
STATE: Indiana  
COUNTRY: United States of America  
ZIP: 46295  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/04294  
FILING DATE:  
CLASSIFICATION:

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/416,257  
FILING DATE: 04-APR-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Blalock, Donna K.  
REGISTRATION NUMBER: 38,082  
REFERENCE/DOCKET NUMBER: X9239  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 317-277-1090  
TELEFAX: 317-276-3861  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 253 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US96-04294-2  
Query Match 100.0%; Score 40; DB 5; Length 253;  
Best Local Similarity 100.0%; Pred. No. 1.8;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 SLLPLQIL 9  
Db 4 SLLPLQIL 12  
RESULT 10  
US-09-502-600-33  
Sequence 33, Application US/09502600A  
GENERAL INFORMATION:  
PATENT NO. 6294344  
APPLICANT: O'Brien, Timothy J.  
TITLE OF INVENTION: Compositions and Methods for the Early Diagnosis of  
Ovarian Cancer  
FILE REFERENCE: D6223CIP-C  
CURRENT FILING DATE: 2000-02-11  
CURRENT APPLICATION NUMBER: US/09/502,600A  
PRIOR APPLICATION NUMBER: 09/039,211  
PRIOR FILING DATE: 03-14-1998  
NUMBER OF SEQ ID NOS: 136  
SEQ ID NO 33  
LENGTH: 9  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: Residues 5-13 of the SCE protein  
US-09-502-600-33  
Query Match 90.0%; Score 36; DB 3; Length 9;  
Best Local Similarity 100.0%; Pred. No. 3e+05;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2 LLLPLQIL 9  
Db 1 LLLPLQIL 8  
RESULT 11  
US-09-918-243-33  
Sequence 33, Application US/09918243  
PATENT NO. 6627403  
GENERAL INFORMATION:  
APPLICANT: O'Brien, Timothy J.  
APPLICANT: Cannon, Martin J.  
APPLICANT: Santin, Alessandro  
TITLE OF INVENTION: Methods for the early diagnosis of ovarian cancer  
FILE REFERENCE: D6223CIP/C/D/CIP  
CURRENT APPLICATION NUMBER: US/09/918,243  
CURRENT FILING DATE: 2001-07-30  
PRIOR APPLICATION NUMBER: US  
PRIOR FILING DATE: 2001-07-13

NUMBER OF SEQ ID NOS: 136

SEQ ID NO 33

LENGTH: 9

TYPE: PRT

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: CHAIN

OTHER INFORMATION: Residues 5-13 of the SCCE protein

US-09-918-243-33

Query Match

Best Local Similarity 90.0%; Score 36; DB 4; Length 9;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 2 LLPLQL 9

Y 1 LLPLQL 8

Y 1 LLPLQL 7

Y 1 LLPLQL 6

Y 1 LLPLQL 5

Y 1 LLPLQL 4

Y 1 LLPLQL 3

Y 1 LLPLQL 2

Y 1 LLPLQL 1

Y 1 LLPLQL 0

Y 1 LLPLQL -1

Y 1 LLPLQL -2

Y 1 LLPLQL -3

Y 1 LLPLQL -4

Y 1 LLPLQL -5

Y 1 LLPLQL -6

Y 1 LLPLQL -7

Y 1 LLPLQL -8

Y 1 LLPLQL -9

Y 1 LLPLQL -10

Y 1 LLPLQL -11

Y 1 LLPLQL -12

Y 1 LLPLQL -13

Y 1 LLPLQL -14

Y 1 LLPLQL -15

Y 1 LLPLQL -16

Y 1 LLPLQL -17

Y 1 LLPLQL -18

Y 1 LLPLQL -19

Y 1 LLPLQL -20

Y 1 LLPLQL -21

Y 1 LLPLQL -22

Y 1 LLPLQL -23

Y 1 LLPLQL -24

Y 1 LLPLQL -25

Y 1 LLPLQL -26

Y 1 LLPLQL -27

Y 1 LLPLQL -28

Y 1 LLPLQL -29

Y 1 LLPLQL -30

Y 1 LLPLQL -31

Y 1 LLPLQL -32

Y 1 LLPLQL -33

Y 1 LLPLQL -34

Y 1 LLPLQL -35

Y 1 LLPLQL -36

Y 1 LLPLQL -37

Y 1 LLPLQL -38

Y 1 LLPLQL -39

Y 1 LLPLQL -40

Y 1 LLPLQL -41

Y 1 LLPLQL -42

Y 1 LLPLQL -43

Y 1 LLPLQL -44

Y 1 LLPLQL -45

Y 1 LLPLQL -46

Y 1 LLPLQL -47

Y 1 LLPLQL -48

Y 1 LLPLQL -49

Y 1 LLPLQL -50

Y 1 LLPLQL -51

Y 1 LLPLQL -52

Y 1 LLPLQL -53

Y 1 LLPLQL -54

Y 1 LLPLQL -55

Y 1 LLPLQL -56

Y 1 LLPLQL -57

Y 1 LLPLQL -58

Y 1 LLPLQL -59

Y 1 LLPLQL -60

Y 1 LLPLQL -61

Y 1 LLPLQL -62

Y 1 LLPLQL -63

Y 1 LLPLQL -64

Y 1 LLPLQL -65

Y 1 LLPLQL -66

Y 1 LLPLQL -67

Y 1 LLPLQL -68

Y 1 LLPLQL -69

Y 1 LLPLQL -70

Y 1 LLPLQL -71

Y 1 LLPLQL -72

Y 1 LLPLQL -73

Y 1 LLPLQL -74

Y 1 LLPLQL -75

Y 1 LLPLQL -76

Y 1 LLPLQL -77

Y 1 LLPLQL -78

Y 1 LLPLQL -79

Y 1 LLPLQL -80

Y 1 LLPLQL -81

Y 1 LLPLQL -82

Y 1 LLPLQL -83

Y 1 LLPLQL -84

Y 1 LLPLQL -85

Y 1 LLPLQL -86

Y 1 LLPLQL -87

Y 1 LLPLQL -88

Y 1 LLPLQL -89

Y 1 LLPLQL -90

Y 1 LLPLQL -91

Y 1 LLPLQL -92

Y 1 LLPLQL -93

Y 1 LLPLQL -94

Y 1 LLPLQL -95

Y 1 LLPLQL -96

Y 1 LLPLQL -97

Y 1 LLPLQL -98

Y 1 LLPLQL -99

Y 1 LLPLQL -100

Y 1 LLPLQL -101

Y 1 LLPLQL -102

Y 1 LLPLQL -103

Y 1 LLPLQL -104

Y 1 LLPLQL -105

Y 1 LLPLQL -106

Y 1 LLPLQL -107

Y 1 LLPLQL -108

Y 1 LLPLQL -109

Y 1 LLPLQL -110

Y 1 LLPLQL -111

Y 1 LLPLQL -112

Y 1 LLPLQL -113

Y 1 LLPLQL -114

Y 1 LLPLQL -115

Y 1 LLPLQL -116

Y 1 LLPLQL -117

Y 1 LLPLQL -118

Y 1 LLPLQL -119

Y 1 LLPLQL -120

Y 1 LLPLQL -121

Y 1 LLPLQL -122

Y 1 LLPLQL -123

Y 1 LLPLQL -124

Y 1 LLPLQL -125

Y 1 LLPLQL -126

Y 1 LLPLQL -127

Y 1 LLPLQL -128

Y 1 LLPLQL -129

Y 1 LLPLQL -130

Y 1 LLPLQL -131

Y 1 LLPLQL -132

Y 1 LLPLQL -133

Y 1 LLPLQL -134

Y 1 LLPLQL -135

Y 1 LLPLQL -136

Y 1 LLPLQL -137

Y 1 LLPLQL -138

Y 1 LLPLQL -139

Y 1 LLPLQL -140

Y 1 LLPLQL -141

Y 1 LLPLQL -142

Y 1 LLPLQL -143

Y 1 LLPLQL -144

Y 1 LLPLQL -145

Y 1 LLPLQL -146

Y 1 LLPLQL -147

Y 1 LLPLQL -148

Y 1 LLPLQL -149

Y 1 LLPLQL -150

Y 1 LLPLQL -151

Y 1 LLPLQL -152

Y 1 LLPLQL -153

Y 1 LLPLQL -154

Y 1 LLPLQL -155

Y 1 LLPLQL -156

Y 1 LLPLQL -157

Y 1 LLPLQL -158

Y 1 LLPLQL -159

Y 1 LLPLQL -160

Y 1 LLPLQL -161

Y 1 LLPLQL -162

Y 1 LLPLQL -163

Y 1 LLPLQL -164

Y 1 LLPLQL -165

Y 1 LLPLQL -166

Y 1 LLPLQL -167

Y 1 LLPLQL -168

Y 1 LLPLQL -169

Y 1 LLPLQL -170

Y 1 LLPLQL -171

Y 1 LLPLQL -172

Y 1 LLPLQL -173

Y 1 LLPLQL -174

Y 1 LLPLQL -175

Y 1 LLPLQL -176

Y 1 LLPLQL -177

Y 1 LLPLQL -178

Y 1 LLPLQL -179

Y 1 LLPLQL -180

Y 1 LLPLQL -181

Y 1 LLPLQL -182

Y 1 LLPLQL -183

Y 1 LLPLQL -184

Y 1 LLPLQL -185

Y 1 LLPLQL -186

Y 1 LLPLQL -187

Y 1 LLPLQL -188

Y 1 LLPLQL -189

Y 1 LLPLQL -190

Y 1 LLPLQL -191

Y 1 LLPLQL -192

Y 1 LLPLQL -193

Y 1 LLPLQL -194

Y 1 LLPLQL -195

Y 1 LLPLQL -196

Y 1 LLPLQL -197

Y 1 LLPLQL -198

Y 1 LLPLQL -199

Y 1 LLPLQL -200

Y 1 LLPLQL -201

Y 1 LLPLQL -202

Y 1 LLPLQL -203

Y 1 LLPLQL -204

Y 1 LLPLQL -205

Y 1 LLPLQL -206

Y 1 LLPLQL -207

Y 1 LLPLQL -208

Y 1 LLPLQL -209

Y 1 LLPLQL -210

Y 1 LLPLQL -211

Y 1 LLPLQL -212

Y 1 LLPLQL -213

Y 1 LLPLQL -214

Y 1 LLPLQL -215

Y 1 LLPLQL -216

Y 1 LLPLQL -217

Y 1 LLPLQL -218

Y 1 LLPLQL -219

Y 1 LLPLQL -220

Y 1 LLPLQL -221

Y 1 LLPLQL -222

Y 1 LLPLQL -223

Y 1 LLPLQL -224

Y 1 LLPLQL -225

Y 1 LLPLQL -226

Y 1 LLPLQL -227

Y 1 LLPLQL -228

Y 1 LLPLQL -229

Y 1 LLPLQL -230

Y 1 LLPLQL -231

Y 1 LLPLQL -232

Y 1 LLPLQL -233

Y 1 LLPLQL -234

Y 1 LLPLQL -235

Y 1 LLPLQL -236

Y 1 LLPLQL -237

Y 1 LLPLQL -238

Y 1 LLPLQL -239

Y 1 LLPLQL -240

Y 1 LLPLQL -241

Y 1 LLPLQL -242

Y 1 LLPLQL -243

Y 1 LLPLQL -244

Y 1 LLPLQL -245

Y 1 LLPLQL -246

Y 1 LLPLQL -247

Y 1 LLPLQL -248

Y 1 LLPLQL -249

Y 1 LLPLQL -250

Y 1 LLPLQL -251

Y 1 LLPLQL -252

Y 1 LLPLQL -253

Y 1 LLPLQL -254

Y 1 LLPLQL -255

Y 1 LLPLQL -256

Y 1 LLPLQL -257

Y 1 LLPLQL -258

Y 1 LLPLQL -259

Y 1 LLPLQL -260

Y 1 LL

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

M protein - protein search, using sw model

un on: March 1, 2004, 17:35:01 ; Search time 24.1111 Seconds  
(without alignments)  
78.818 Million cell updates/sec

itle: US-09-905-083-36  
erfect score: 40  
equence: 1 SLLPLQL 9

coring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

earched: 809742 seqs, 211153259 residues

otal number of hits satisfying chosen parameters: 809742

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

ost-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

atabase : Published Applications AA.\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US03\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/1/pubpaa/FCUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	40	100.0	9	9	US-09-918-243-36
2	40	100.0	9	9	Sequence 36, Appl
3	40	100.0	253	9	US-09-905-083-36
4	40	100.0	253	9	Sequence 98, Appl
5	40	100.0	253	9	Sequence 3, Appl
6	40	100.0	253	14	US-10-264-283-90
7	40	100.0	253	15	US-10-295-027-498
8	36	90.0	9	9	US-10-173-999-48
9	36	90.0	9	9	US-09-918-243-33
10	35	87.5	868	15	US-10-369-493-10255
11	33	82.5	804	14	US-10-236-055A-28
12	32	80.0	9	9	US-09-918-243-35
13	32	80.0	9	9	US-09-918-243-116
14	32	80.0	9	9	US-09-905-083-35
15	32	80.0	9	9	US-09-905-083-116

16	32	80.0	139	15	US-10-104-047-3867
17	32	80.0	282	15	US-10-302-267-148
18	32	80.0	282	15	US-10-374-780A-2572
19	32	80.0	559	14	US-10-156-761-10881
20	31	77.5	46	14	US-10-050-882-135
21	31	77.5	211	14	US-10-050-882-131
22	31	77.5	246	10	US-09-948-374-338
23	31	77.5	246	14	US-10-008-856A-338
24	31	77.5	246	14	US-10-008-818A-338
25	31	77.5	246	14	US-10-015-393A-338
26	31	77.5	246	14	US-10-015-869A-338
27	31	77.5	246	14	US-10-012-121A-338
28	31	77.5	246	14	US-10-008-116A-338
29	31	77.5	246	14	US-10-006-117A-338
30	31	77.5	246	14	US-10-017-527A-338
31	31	77.5	246	14	US-10-013-913A-338
32	31	77.5	246	14	US-10-007-194A-338
33	31	77.5	246	14	US-10-013-430A-338
34	31	77.5	246	14	US-10-011-671A-338
35	31	77.5	246	14	US-10-012-755A-338
36	31	77.5	246	14	US-10-015-386A-338
37	31	77.5	246	14	US-10-011-692A-338
38	31	77.5	246	14	US-10-006-768A-338
39	31	77.5	246	14	US-10-017-610A-338
40	31	77.5	246	14	US-10-006-063A-338
41	31	77.5	246	14	US-10-020-063A-338
42	31	77.5	246	14	US-10-015-391A-338
43	31	77.5	246	14	US-10-017-407A-338
44	31	77.5	246	14	US-10-011-833A-338
45	31	77.5	246	14	US-10-006-041A-338

ALIGNMENTS

RESULT 1

US-09-918-243-36  
; Sequence 36, Application US/09918243  
; Patent No. US20020142317A1  
; GENERAL INFORMATION:  
; APPLICANT: O'Brien, Timothy J.  
; APPLICANT: Cannon, Martin J.  
; APPLICANT: Santin, Alessandro  
; TITLE OF INVENTION: Methods for the early diagnosis of ovarian cancer  
; FILE REFERENCE: D6223CIP/C/D/CIP  
; CURRENT APPLICATION NUMBER: US/09/918,243  
; CURRENT FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US  
; PRIOR FILING DATE: 2001-07-13  
; NUMBER OF SEQ ID NOS: 136  
; SEQ ID NO 36  
; LENGTH: 9  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CHAIN  
; OTHER INFORMATION: Residues 4-12 of the SCCE protein  
US-09-918-243-36

Query Match 100.0%; Score 40; DB 9; Length 9;  
Best Local Similarity 100.0%; Pred. No. 7.1e+05;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SLLPLQL 9

Db 1 SLLPLQL 9

RESULT 2

US-09-905-083-36  
; Sequence 36, Application US/09905083  
; Patent No. US20020146708A1  
; GENERAL INFORMATION:

APPLICANT: O'Brien, Timothy J.  
TITLE OF INVENTION: Compositions and Methods for the Early Diagnosis of  
TITLE OF INVENTION: Ovarian Cancer  
FILE REFERENCE: D6223CIP/C/DIV  
CURRENT APPLICATION NUMBER: US/09/905,083  
CURRENT FILING DATE: 2001-07-13  
PRIOR APPLICATION NUMBER: US 09/502,600  
PRIOR FILING DATE: 2000-02-11  
NUMBER OF SEQ ID NOS: 136  
SEQ ID NO 36  
LENGTH: 9  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CHAIN  
OTHER INFORMATION: Residues 4-12 of the SCE protein  
S-09-905-083-36  
Query Match 100.0%; Score 40; DB 9; Length 9;  
Best Local Similarity 100.0%; Pred. No. 7.1e+05;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Y 1 SLLPLQL 9  
| | | | |  
b 1 SLLPLQL 9  
| | | | |  
RESULT 3  
S-09-888-615-98  
Sequence 98, Application US/09888615  
Patent No. US20020064856A1  
GENERAL INFORMATION:  
APPLICANT: PLOWMAN, GREGORY  
APPLICANT: WHYTE, DAVID  
APPLICANT: CAENEPEEL, SEAN  
APPLICANT: CHARYDCZAK, GLEN  
APPLICANT: MANNING, GERARD  
APPLICANT: SUDARSANAM, SUCHA  
TITLE OF INVENTION: NOVEL PROTEASES  
FILE REFERENCE: 038602/1214  
CURRENT APPLICATION NUMBER: US/09/888,615  
CURRENT FILING DATE: 2001-06-26  
PRIOR APPLICATION NUMBER: 60/214,047  
PRIOR FILING DATE: 2000-06-26  
NUMBER OF SEQ ID NOS: 150  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 98  
LENGTH: 253  
TYPE: PRT  
ORGANISM: Homo sapiens  
IS-09-888-615-98  
Query Match 100.0%; Score 40; DB 9; Length 253;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Y 1 SLLPLQL 9  
| | | | |  
b 4 SLLPLQL 12  
| | | | |  
RESULT 4  
IS-09-764-762-3  
Sequence 3, Application US/09764762  
Patent No. US20020068341A1  
GENERAL INFORMATION:  
APPLICANT: Hillman, Jennifer L.  
Lal, Preeti  
TITLE OF INVENTION: NOVEL KALLIKREIN  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive

CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/764,762  
FILING DATE: 16-Jan-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/210,084  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0252 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 253 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 532504  
SEQUENCE DESCRIPTION: SEQ ID NO: 3:  
US-09-764-762-3  
Query Match 100.0%; Score 40; DB 9; Length 253;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Y 1 SLLPLQL 9  
| | | | |  
b 4 SLLPLQL 12  
| | | | |  
RESULT 5  
US-10-264-283-90  
Sequence 90, Application US/10264283  
Publication No. US2003014494A1  
GENERAL INFORMATION:  
APPLICANT: Algate, Paul A.  
APPLICANT: Mannion, Jane  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER  
FILE REFERENCE: 210121.590  
CURRENT APPLICATION NUMBER: US/10/264,283  
CURRENT FILING DATE: 2002-10-02  
NUMBER OF SEQ ID NOS: 111  
SOFTWARE: Corixa Invention Disclosure Database  
SEQ ID NO 90  
LENGTH: 253  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-264-283-90  
Query Match 100.0%; Score 40; DB 14; Length 253;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Y 1 SLLPLQL 9  
| | | | |  
b 4 SLLPLQL 12  
| | | | |

```

RESULT 6
US-10-295-027-498
Sequence 498, Application US/10295027
Publication No. US20030232350A1
GENERAL INFORMATION:
APPLICANT: Afar, Daniel
APPLICANT: Aziz, Natasha
APPLICANT: Ginsberg, Wendy M.
APPLICANT: Gish, Kurt C.
APPLICANT: Glynn, Richard
APPLICANT: Hevezi, Peter A.
APPLICANT: Mack, David H.
APPLICANT: Murray, Richard
APPLICANT: Watson, Susan R.
APPLICANT: Eos Biotechnology, Inc.
TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
FILE REFERENCE: 018501-012500US
CURRENT FILING DATE: 2002-11-13
PRIOR APPLICATION NUMBER: US 09/663,733
PRIOR FILING DATE: 2000-09-15
PRIOR APPLICATION NUMBER: US 60/350,666
PRIOR FILING DATE: 2001-11-13
PRIOR APPLICATION NUMBER: US 60/335,394
PRIOR FILING DATE: 2001-11-15
PRIOR APPLICATION NUMBER: US 60/332,464
PRIOR FILING DATE: 2001-11-21
PRIOR APPLICATION NUMBER: US 60/334,393
PRIOR FILING DATE: 2001-11-29
PRIOR APPLICATION NUMBER: US 60/340,376
PRIOR FILING DATE: 2001-12-14
PRIOR APPLICATION NUMBER: US 60/347,211
PRIOR FILING DATE: 2002-01-08
PRIOR APPLICATION NUMBER: US 60/347,349
PRIOR FILING DATE: 2002-01-10
PRIOR APPLICATION NUMBER: US 60/355,250
PRIOR FILING DATE: 2002-02-08
PRIOR APPLICATION NUMBER: US 60/356,714
PRIOR FILING DATE: 2002-02-13
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 1386
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 498
LENGTH: 253
TYPE: PRT
ORGANISM: Homo sapiens
US-10-295-027-498

Query Match 100.0%; Score 40; DB 15; Length 253;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2y 1 SLLPLQIL 9
   |||||
Db 4 SLLPLQIL 12

RESULT 7
US-10-173-999-48
Sequence 48, Application US/10173999
Publication No. US20040005563A1
GENERAL INFORMATION:
APPLICANT: Mack, David H.
APPLICANT: Gish, Kurt C.
APPLICANT: Eos Biotechnology, Inc.
TITLE OF INVENTION: Methods of Diagnosis of Ovarian Cancer, Compositions
FILE REFERENCE: 018501-002420US
CURRENT FILING DATE: 2002-06-17
PRIOR APPLICATION NUMBER: US/10/173,999
PRIOR FILING DATE: 2002-06-17
PRIOR APPLICATION NUMBER: US 60/299,234

```

```

; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: US 60/315,287
; PRIOR FILING DATE: 2001-08-27
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/372,246
; PRIOR FILING DATE: 2001-04-12
; NUMBER OF SEQ ID NOS: 163
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 48
; LENGTH: 253
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-173-999-48

Query Match 100.0%; Score 40; DB 15; Length 253;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SLLPLQIL 9
   |||||
Db 4 SLLPLQIL 12

RESULT 8
US-09-918-243-33
Sequence 33, Application US/09918243
Patent No. US20020142317A1
GENERAL INFORMATION:
APPLICANT: O'Brien, Timothy J.
APPLICANT: Santin, Alessandro
TITLE OF INVENTION: Methods for the early diagnosis of ovarian cancer
FILE REFERENCE: D6223CIP/C/D/CIP
CURRENT APPLICATION NUMBER: US/09/918,243
CURRENT FILING DATE: 2001-07-30
PRIOR APPLICATION NUMBER: US
PRIOR FILING DATE: 2001-07-13
NUMBER OF SEQ ID NOS: 136
SEQ ID NO 33
LENGTH: 9
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CHAIN
OTHER INFORMATION: Residues 5-13 of the SCCE protein
US-09-918-243-33

Query Match 90.0%; Score 36; DB 9; Length 9;
Best Local Similarity 100.0%; Pred. No. 7;ie+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LLLPLQIL 9
   |||||
Db 1 LLLPLQIL 8

RESULT 9
US-09-905-083-33
Sequence 33, Application US/09905083
Patent No. US20020146708A1
GENERAL INFORMATION:
APPLICANT: O'Brien, Timothy J.
TITLE OF INVENTION: Compositions and Methods for the Early Diagnosis of
FILE REFERENCE: D6223CIP/C/Div
CURRENT APPLICATION NUMBER: US/09/905,083
CURRENT FILING DATE: 2001-07-13
PRIOR APPLICATION NUMBER: US 09/502,600
PRIOR FILING DATE: 2000-02-11
NUMBER OF SEQ ID NOS: 136
SEQ ID NO 33
LENGTH: 9

```

```

TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CHAIN
OTHER INFORMATION: Residues 5-13 of the SCCE protein
5-09-905-083-33

Query Match      90.0%; Score 36; DB 9; Length 9;
Best Local Similarity 100.0%; Pred. No. 7.1e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 2 LLPLQL 9
  |||||
b 1 LLPLQL 8

RESULT 10
5-10-369-493-10255
Sequence 10255, Application US/10369493
Publication No. US2003033675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 10255
LENGTH: 868
TYPE: PRT
ORGANISM: Cytophaga hutchinsonii
5-10-369-493-10255

Query Match      87.5%; Score 35; DB 15; Length 868;
Best Local Similarity 88.9%; Pred. No. 3.8e+02;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Y 1 SLLPLQL 9
  |||||
b 702 SLLPLQL 710

RESULT 11
5-10-236-055A-28
Sequence 28, Application US/10236055A
Publication No. US20030134328A1
GENERAL INFORMATION:
APPLICANT: Basham, Beth E.
APPLICANT: Forsythe, Ian
APPLICANT: Gorman, Daniel M.
APPLICANT: Matteson, Jeanine
APPLICANT: Moshrefi, Mehrdad
APPLICANT: Parham, Christi
TITLE OF INVENTION: MAMMALIAN GENES; RELATED REAGENTS
FILE REFERENCE: EX01343X
CURRENT APPLICATION NUMBER: US/10/236,055A
CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/317,988
PRIOR FILING DATE: 2001-09-06
NUMBER OF SEQ ID NOS: 36
SOFTWARE: Patentin version 3.1
SEQ ID NO 28
LENGTH: 804
TYPE: PRT
ORGANISM: Mus musculus
5-10-236-055A-28

```

```

Query Match      82.5%; Score 33; DB 14; Length 804;
Best Local Similarity 75.0%; Pred. No. 8e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SLLPLQL 8
  |||||
Db 760 SLLPLQL 767

RESULT 12
US-09-918-243-35
Sequence 35, Application US/09918243
Patent No. US20020142317A1
GENERAL INFORMATION:
APPLICANT: O'Brien, Timothy J.
APPLICANT: Cannon, Martin J.
APPLICANT: Santin, Alessandro
TITLE OF INVENTION: Methods for the early diagnosis of ovarian cancer
FILE REFERENCE: D6223CIP/C/D/CIP
CURRENT APPLICATION NUMBER: US/09/918,243
CURRENT FILING DATE: 2001-07-30
PRIOR APPLICATION NUMBER: US
PRIOR FILING DATE: 2001-07-13
NUMBER OF SEQ ID NOS: 136
SEQ ID NO 35
LENGTH: 9
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CHAIN
OTHER INFORMATION: Residues 6-14 of the SCCE protein
US-09-918-243-35

Query Match      80.0%; Score 32; DB 9; Length 9;
Best Local Similarity 100.0%; Pred. No. 7.1e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 LLPLQL 9
  |||||
Db 1 LLPLQL 7

RESULT 13
US-09-918-243-116
Sequence 116, Application US/09918243
Patent No. US20020142317A1
GENERAL INFORMATION:
APPLICANT: O'Brien, Timothy J.
APPLICANT: Cannon, Martin J.
APPLICANT: Santin, Alessandro
TITLE OF INVENTION: Methods for the early diagnosis of ovarian cancer
FILE REFERENCE: D6223CIP/C/D/CIP
CURRENT APPLICATION NUMBER: US/09/918,243
CURRENT FILING DATE: 2001-07-30
PRIOR APPLICATION NUMBER: US
PRIOR FILING DATE: 2001-07-13
NUMBER OF SEQ ID NOS: 136
SEQ ID NO 116
LENGTH: 9
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CHAIN
OTHER INFORMATION: Residues 2-10 of the SCCE protein
US-09-918-243-116

Query Match      80.0%; Score 32; DB 9; Length 9;
Best Local Similarity 100.0%; Pred. No. 7.1e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SLLPLQL 7
  |||||

```

db 3 SLLPLQ 9

## RESULT 14

US-09-905-083-35  
 ; Sequence 35, Application US/09905083  
 ; Patent No. US20020146708A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: O'Brien, Timothy J.  
 ; TITLE OF INVENTION: Compositions and Methods for the Early Diagnosis of  
 ; FILE REFERENCE: D6223CIP/C/Div  
 ; CURRENT APPLICATION NUMBER: US/09/905,083  
 ; CURRENT FILING DATE: 2001-07-13  
 ; PRIOR APPLICATION NUMBER: US 09/502,600  
 ; PRIOR FILING DATE: 2000-02-11  
 ; NUMBER OF SEQ ID NOS: 136  
 ; SEQ ID NO 35  
 ; LENGTH: 9  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: CHAIN  
 ; OTHER INFORMATION: Residues 6-14 of the SCCE protein  
 ; US-09-905-083-35

Query Match 80.0%; Score 32; DB 9; Length 9;  
 Best Local Similarity 100.0%; Pred. No. 7.1e+05;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2y 3 LLPLQL 9  
 |||||  
 Db 1 LLPLQL 7

## RESULT 15

US-09-905-083-116  
 ; Sequence 116, Application US/09905083  
 ; Patent No. US20020146708A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: O'Brien, Timothy J.  
 ; TITLE OF INVENTION: Compositions and Methods for the Early Diagnosis of  
 ; FILE REFERENCE: D6223CIP/C/Div  
 ; CURRENT APPLICATION NUMBER: US/09/905,083  
 ; CURRENT FILING DATE: 2001-07-13  
 ; PRIOR APPLICATION NUMBER: US 09/502,600  
 ; PRIOR FILING DATE: 2000-02-11  
 ; NUMBER OF SEQ ID NOS: 136  
 ; SEQ ID NO 116  
 ; LENGTH: 9  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: CHAIN  
 ; OTHER INFORMATION: Residues 2-10 of the SCCE protein  
 ; US-09-905-083-116

Query Match 80.0%; Score 32; DB 9; Length 9;  
 Best Local Similarity 100.0%; Pred. No. 7.1e+05;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SLLPLQ 7  
 |||||  
 Db 3 SLLPLQ 9

Search completed: March 1, 2004, 18:08:52  
 Job time : 25.1111 secs



GenCore version 5.1.6  
Copyright (C) 1993 - 2004 Compugen Ltd.

M protein - protein search, using sw model

un on: March 1, 2004, 17:21:01 ; Search time 10.2222 Seconds  
(without alignments)  
84.690 Million cell updates/sec

itle: US-09-905-083-36

effect score: 40

equences: 1 SLLPLQL 9

coring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

searched: 283366 seqs, 96191526 residues

total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR 78:\*\*

2: PIR2:\*\*

3: PIR3:\*\*

4: PIR4:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

result No.	Score	Query Match	Length	DB ID	Description
1	40	100.0	253	A53968	serine proteinase
2	38	95.0	146	H75201	hypothetical prote
3	33	82.5	370	AB3334	daunorubicin resis
4	32	80.0	218	T02912	probable transcrip
5	32	80.0	264	AC2620	1-acyl-sn-glycerol
6	32	80.0	264	C97402	probable acyltrans
7	32	80.0	266	JC7300	tax-responsive ele
8	32	80.0	267	JC4857	hepatocarcinogenes
9	32	80.0	282	B85327	probable transcrip
10	32	80.0	494	C49349	probable succinogl
11	32	80.0	494	E95976	probable transport
12	32	80.0	582	S40176	Exor protein - Rhl
13	32	80.0	622	S61692	probable membrane
14	31	77.5	196	G65039	hypothetical prote
15	31	77.5	303	C84914	hypothetical prote
16	31	77.5	354	T48649	glycerol-3-phospha
17	31	77.5	372	A98157	probable permease
18	31	77.5	372	AH3130	ABC transporter, m
19	31	77.5	392	D83934	hypothetical prote
20	31	77.5	395	D69779	antibiotic resista
21	31	77.5	398	C91063	hypothetical prote
22	31	77.5	401	B90120	SNF1-related prote
23	31	77.5	413	A20834	probable membrane
24	31	77.5	470	A90083	hypothetical prote
25	31	77.5	475	A69149	O-antigen transpor
26	31	77.5	754	A60614	probable competenc
27	31	77.5	783	A46136	Myosin-heavy-chain
28	31	77.5	913	AC2445	hypothetical prote
29	30	75.0	155	S38662	interleukin-2 - go

interleukin-2 prec  
interleukin-2 prec  
probable membrane  
hypothetical prote  
probable phosphate  
NADH2 dehydrogenas  
hypothetical prote  
cell division prot  
cytochrome a3 con  
probable high affi  
T-cell surface gly  
conserved hypothet  
surfactant protein  
conglutinin precur  
conglutinin - bovi  
probable transport

#### ALIGNMENTS

##### RESULT 1

A53968  
serine proteinase SCCE precursor - human  
N:Alternate names: stratum corneum chymotryptic enzyme  
C:Species: Homo sapiens (man)  
C>Date: 07-Jul-1995 #sequence\_revision 07-Jul-1995 #text\_change 22-Jun-1999  
C:Accession: A53968  
R:Hansson, L.; Stroemqvist, M.; Baeckman, A.; Wallbrandt, P.; Carlstein, A.; Egelrud, T.  
J. Biol. Chem. 269, 19420-19426, 1994  
A:Title: Cloning, expression, and characterization of stratum corneum chymotryptic enzym  
A:Reference number: A53968; MUID:94308225; PMID:8034709  
A:Accession: A53968  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-253 <HAN>  
A:Cross-references: GB:L33404; NID:G521214; PIDN:AAC37551.1; PID:G532504  
C:Genetics:  
A:Gene: GDB:PRSS6; SCCE  
A:Cross-references: GDB:377730  
A:Map position: 7q35-7q35  
C:Superfamily: trypsin; trypsin homology  
F:30-245/Domain: trypsin homology <TRY>

Query Match 100.0%; Score 40; DB 2; Length 253;  
Best Local Similarity 100.0%; Pred. No. 1.5;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SLLPLQL 9  
DB 4 SLLPLQL 12

##### RESULT 2

H75201  
hypothetical protein PAB0088 - Pyrococcus abyssi (strain Orsay)  
C:Species: Pyrococcus abyssi  
C>Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 08-Sep-2000  
C:Accession: H75201  
R:anonymous, Genoscope  
submitted to the EMBL Data Library, July 1999  
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome stru  
A:Reference number: A75001  
A:Accession: H75201  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-146 <KAW>  
A:Cross-references: GB:AJ248283; GB:AL096836; NID:G5457433; PIDN:CAB49063.1; PID:el51495  
A:Experimental source: strain Orsay  
C:Genetics:  
A:Gene: PAB0088  
C:Superfamily: Pyrococcus abyssi hypothetical protein PAB0088

Query Match 95.0%; Score 38; DB 2; Length 146;  
 Best Local Similarity 88.9%; Pred. No. 2,2; Mismatches 1; Gaps 0;  
 Matches 8; Conservative 1; Indels 0;

Y 1 SLLPLQIL 9  
 :|||||:  
 b 66 SLLPLQIL 74

RESULT 3  
 B3334  
 anorubicin resistance transmembrane protein [imported] - Brucella melitensis (strain 1  
 ;Species: Brucella melitensis  
 ;Date: 01-Feb-2002 #sequence\_revision 01-Feb-2002 #text\_change 01-Feb-2002  
 ;Accession: AB3334  
 ;DelVecchio, V.G.; Kaputal, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova,  
 ;Mazur, M.; Goltzman, E.; Selkov, E.; Elizer, P.H.; Hagius, S.; O'Callaghan, D.; Letess  
 ;roc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002  
 ;Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis  
 ;Reference number: AD3252; PMID:11756888  
 ;Accession: AB3334  
 ;Status: preliminary  
 ;Molecule type: DNA  
 ;Residues: 1-370 <KUR>  
 ;Cross-references: GB:AE008917; PIDN:AAL51837.1; PID:gi7982584; GSPDB:GN00190  
 ;Experimental source: strain 16M  
 ;Genetics:  
 ;Gene: BMEI0656  
 ;Map position: 1

Query Match 82.5%; Score 33; DB 2; Length 370;  
 Best Local Similarity 75.0%; Pred. No. 57; Mismatches 2; Gaps 0;  
 Matches 6; Conservative 2; Indels 0;

Y 2 LLLPLQIL 9  
 :|||||:  
 b 293 ILLPLQVL 300

RESULT 4  
 02912  
 probable transcription factor MYB41 [similarity] - Arabidopsis thaliana (fragment)  
 ;Alternate names: protein T13J8.220  
 ;Species: Arabidopsis thaliana (mouse-ear cress)  
 ;Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 02-Sep-2000  
 ;Accession: T02912; T51654  
 ;Bevan, M.; Pohl, T.; Weizenegger, T.; Hoheisel, J.; Mewes, H.W.; Mayer, K.F.X.; Schuel  
 ;submitted to the Protein Sequence Database, February 1999  
 ;Reference number: Z14766  
 ;Accession: T02912  
 ;Molecule type: DNA  
 ;Residues: 1-218 <BEV>  
 ;Cross-references: EMBL:AL035524  
 ;Experimental source: cultivar Columbia; BAC clone T13J8  
 ;Kranz, H.D.; Denekamp, M.; Greco, R.; Jin, H.; Kranz, H.D.; Denekamp, M.; Greco, R.; J  
 ;Paz-Ares, J.; Weishaar, B.  
 ;Plant J. 16, 263-276, 1998  
 ;Title: Towards functional characterisation of the members of the R2R3-MYB gene from Ar  
 ;Reference number: Z14349; MUID:9839469; PMID:9839469  
 ;Accession: T51654  
 ;Status: preliminary; translated from GB/EMBL/DBJ  
 ;Molecule type: mRNA  
 ;Residues: 1-218 <KRA>  
 ;Cross-references: EMBL:AF062882; PIDN:AAC63604.1  
 ;Experimental source: cultivar Columbia  
 ;Genetics:  
 ;Gene: MYB41  
 ;Map position: 4  
 ;Note: T13J8.220  
 ;Superfamily: unassigned myb DNA-binding repeat proteins; myb DNA-binding repeat homolo  
 ;Keywords: transcription factor  
 ;1-48/Domain: myb DNA-binding repeat homology (fragment) <MYB>

Query Match 80.0%; Score 32; DB 2; Length 218;  
 Best Local Similarity 100.0%; Pred. No. 52; Mismatches 0; Gaps 0;  
 Matches 7; Conservative 0; Indels 0;

QY 1 SLLPLQ 7  
 :|||||:  
 Db 104 SLLPLQ 110

RESULT 5  
 AC2820  
 1-acyl-sn-glycerol-3-phosphate acyltransferase plsC [imported] - Agrobacterium tumefaciens  
 ;Species: Agrobacterium tumefaciens  
 ;Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 18-Nov-2002  
 ;Accession: AC2620  
 ;R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.  
 ;erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell  
 ;Karp, P.; Romero, P.; Zhang, S.  
 ;Science 294, 2317-2323, 2001  
 ;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, E  
 ;ster, E.W.  
 ;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
 ;Reference number: AB2577; MUID:21608550; PMID:11743193  
 ;Accession: AC2620  
 ;Status: preliminary  
 ;Molecule type: DNA  
 ;Residues: 1-264 <KUR>  
 ;Cross-references: GB:AE008688; PIDN:AAL41377.1; PID:gi7738693; GSPDB:GN00186  
 ;Experimental source: strain C58 (Dupont)  
 ;Genetics:  
 ;Gene: plsC  
 ;Map position: circular chromosome

Query Match 80.0%; Score 32; DB 2; Length 264;  
 Best Local Similarity 75.0%; Pred. No. 64; Mismatches 2; Indels 0; Gaps 0;  
 Matches 6; Conservative 2; Indels 0;

QY 2 LLLPLQIL 9  
 :|||||:  
 Db 19 ILLPLQIL 26

RESULT 6  
 C97402  
 probable acyltransferase (AF232919) [imported] - Agrobacterium tumefaciens (strain C58, C  
 ;Species: Agrobacterium tumefaciens  
 ;Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 18-Nov-2002  
 ;Accession: C97402  
 ;R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorollo, B.; Goldman,  
 ;A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;  
 ;Science 294, 2323-2328, 2001  
 ;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tume  
 ;Reference number: A97359; MUID:21608551; PMID:11743194  
 ;Accession: C97402  
 ;Status: preliminary  
 ;Molecule type: DNA  
 ;Residues: 1-264 <KUR>  
 ;Cross-references: GB:AE007869; PIDN:AAK86172.1; PID:gi15155265; GSPDB:GN00169  
 ;Genetics:  
 ;Gene: AGR C 621  
 ;Map position: circular chromosome

Query Match 80.0%; Score 32; DB 2; Length 264;  
 Best Local Similarity 75.0%; Pred. No. 64; Mismatches 2; Indels 0; Gaps 0;  
 Matches 6; Conservative 2; Indels 0;

QY 2 LLLPLQIL 9  
 :|||||:  
 Db 19 ILLPLQIL 26

RESULT 7

```

C7300
ax-responsive element-binding protein 5 - mouse
;Species: Mus musculus (house mouse)
;Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 02-Sep-2000
;Accession: JC7300
;Masaki, T.; Noguchi, H.; Kobayashi, M.; Yoshida, M.; Takamatsu, K.
NA Res. 7, 187-193, 2000
;Title: Isolation and characterization of the gene encoding mouse tax-responsive element
;Reference number: JC7300
;Accession: JC7300
;Molecule type: mRNA
;Residues: 1-266 <MAS>
;Cross-references: DDBJ:AB036745
;Genetics:
;Gene: treb5
;Superfamily: X box-binding protein 1; fos/jun DNA-binding domain homology
;Keywords: leucine zipper; transcription factor

Query Match      80.0%; Score 32; DB 2; Length 266;
Best Local Similarity 100.0%; Pred. No. 64;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y      3 LLPLQL 9
      |||||
b     187 LLPLQL 193

RESULT 8
C4857
epitocarcinogenesis-related transcription factor - rat
;Species: Rattus norvegicus (Norway rat)
;Date: 15-Aug-1996 #sequence_revision 18-Oct-1996 #text_change 08-Sep-1997
;Accession: JC4857
;Kishimoto, T.; Kokura, K.; Kumagai, Y.; Makino, Y.; Tamura, T.
J. Biochem. Biophys. Res. Commun. 224, 746-751, 1996
;Title: HTF: A b-zip transcription factor that is closely related to the human XBP/TRF
;Reference number: JC4857
;Accession: JC4857
;Molecule type: mRNA
;Residues: 1-267 <XIS>
;Comment: This is a basic-leucine zipper type transcription factor involved in hepatoc
;Genetics:
;Gene: htf
;Superfamily: X box-binding protein 1; fos/jun DNA-binding domain homology
;Keywords: leucine zipper; transcription factor
;88-96/Domain: fos/jun DNA-binding domain homology <FJD>
;89-126/Region: leucine zipper motif

Query Match      80.0%; Score 32; DB 2; Length 267;
Best Local Similarity 100.0%; Pred. No. 64;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y      3 LLPLQL 9
      |||||
b     188 LLPLQL 194

RESULT 9
B8327
probable transcription factor MYB41 [imported] - Arabidopsis thaliana
;Species: Arabidopsis thaliana (mouse-ear cress)
;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 02-Mar-2001
;Accession: B8327
;anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Sprin
ature 402, 769-777, 1999
;Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.
;Reference number: A85001; MUID:20083488; PMID:10617198
;Accession: B8327
;Status: preliminary
;Molecule type: DNA
;Residues: 1-282 <STO>
;Cross-references: GB:NC_001268; NID:G7269665; PIDN:CAB79613.1; GSPDB:GN00140
;Genetics:

```

```

A:Gene: AT4g28110
A:Map position: 4
C:Superfamily: barley myb-related protein 3; myb DNA-binding repeat homology

Query Match      80.0%; Score 32; DB 2; Length 282;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY     1 SLLPLQ 7
      |||||
Db    168 SLLPLQ 174

RESULT 10
C49349
probable succinoglycan transport protein ExoT - Rhizobium meliloti
C:Species: Rhizobium meliloti
C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 17-Mar-2000
C:Accession: C49349
R:Glucksmann, M.A.; Reuber, T.L.; Walker, G.C.
J. Bacteriol. 175, 7045-7055, 1993
A:Title: Genes needed for the modification, polymerization, export, and processing of su
A:Reference number: A49349; MUID:94042870; PMID:8226646
A:Accession: C49349
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-494 <GLJ>
A:Cross-references: GB:L20758; NID:G393240; PIDN:AAA16050.1; PID:G393249
C:Superfamily: hypothetical protein b2046
C:Keywords: transmembrane protein

Query Match      80.0%; Score 32; DB 2; Length 494;
Best Local Similarity 75.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY     1 SLLPLQ 8
      |||||
Db    122 SLLPLQ 129

RESULT 11
B95976
probable transport protein, similar to Wzx exoT [imported] - Sinorhizobium meliloti (str
C:Species: Sinorhizobium meliloti
C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
C:Accession: B95976
R:Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernan
Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
A:Title: The complete sequence of the 1.683-kb pSymB megaplasmid from the N2-fixing endo
A:Reference number: A95942; MUID:21396508; PMID:11481431
A:Accession: B95976
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-494 <KUR>
A:Cross-references: GB:AL591985; PIDN:CAC49474.1; PID:G15140960; GSPDB:GN00167
A:Experimental source: strain 1021, megaplasmid pSymB
R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,
pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,
hebaull, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
A:Reference number: A96039; MUID:21368234; PMID:11474104
A:Contents: annotation
C:Genetics:
A:Gene: exoT; SM20950
A:Genome: plasmid
C:Superfamily: hypothetical protein b2046

Query Match      80.0%; Score 32; DB 2; Length 494;
Best Local Similarity 75.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```

```

Y      1 SLLPLQI 8
b      122 SLLPLQL 129

RESULT 12
xOT protein - Rhizobium meliloti
Species: Rhizobium meliloti
Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 19-May-2000
Accession: S40176
Becker, A.; Kleickmann, A.; Kuester, H.; Keller, M.; Arnold, W.; Puhler, A.
Submitted to the EMBL Data Library, April 1993
Description: Analysis of the Rhizobium meliloti genes exoU, exoV, exoW, exoT and exoI
es.
Reference number: S40173
Accession: S40176
Status: preliminary
Molecule type: DNA
Residues: 1-582 <BEC>
Cross-references: EMBL:Z22646
Superfamily: hypothetical protein b2046

Query Match      80.0%; Score 32; DB 2; Length 582;
Best Local Similarity 75.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Y      1 SLLPLQI 8
b      122 SLLPLQL 129

RESULT 13
robable membrane protein YOR137c - Yeast (Saccharomyces cerevisiae)
Alternate names: hypothetical protein O3329; hypothetical protein YOR3329c
Species: Saccharomyces cerevisiae
Date: 09-Mar-1996 #sequence_revision 12-Apr-1996 #text_change 19-Apr-2002
Accession: S61692; S67022
Benes, V.; Andrade, M.A.; Rechmann, S.; Teodoru, C.; Banrevi, A.; Sander, C.; Valencia
Submitted to the EMBL Data Library, December 1995
Description: Nucleotide sequence and analysis of a 130 kb fragment of yeast chromosome
Reference number: S61643
Accession: S61692
Molecule type: DNA
Residues: 1-622 <BEN>
Cross-references: EMBL:X94335; NID:gl262139; PID:e217839; PID:gl164980
Voss, H.; Benes, V.; Rechmann, S.; Teodoru, C.; Schwager, C.; Faces, V.; Ansoerge, W.
Submitted to the Protein Sequence Database, July 1996
Reference number: S66965
Accession: S67022
Molecule type: DNA
Residues: 1-622 <VOS>
Cross-references: EMBL:Z75045; NID:gl420348; PID:e252028; PID:gl420349; MIPS:YOR137c
Genetics:
Experimental source: strain S288C
Map position: 15R
Keywords: transmembrane protein
11-27/Domain: transmembrane #status predicted <TM>

Query Match      80.0%; Score 32; DB 2; Length 622;
Best Local Similarity 66.7%; Pred. No. 1.5e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Y      1 SLLPLQL 9
b      141 SALLPLQL 149

RESULT 14
65039

```

```

hypothetical protein b2612 - Escherichia coli (strain K-12)
C;Species: Escherichia coli
C;Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002
C;Accession: G65039
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Col
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617; PMID:9278503
A;Accession: G65039
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-196 <BIAT>
A;Cross-references: GB:AE000347; GB:U00096; NID:G2367142; PID:AACT5661.1; PID:gl788965;
A;Experimental source: strain K-12, substrain MG1655

Query Match      77.5%; Score 31; DB 2; Length 196;
Best Local Similarity 87.5%; Pred. No. 74;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Y      2 LLLPLQL 9
b      123 LLAPLQL 130

RESULT 15
C84914
hypothetical protein At2g47360 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C;Accession: C84914
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.L.; Town, C.D.; Fujii, C.Y.; N
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: C84914
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-303 <STO>
A;Cross-references: GB:AE002093; NID:g2275214; PID:AAE63836.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g47360
A;Map position: 2

Query Match      77.5%; Score 31; DB 2; Length 303;
Best Local Similarity 77.8%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Y      1 SLLPLQL 9
b      29 SLLPLSL 37

Search completed: March 1, 2004, 17:36:30
Job time : 10.3333 secs

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

2M protein - protein search, using sw model

Run on: March 1, 2004, 17:17:25 ; Search time 6 Seconds  
(without alignments)  
78.105 Million cell updates/sec

Title: US-09-905-083-36

Perfect score: 40

Sequence: 1 SLLPLQL 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

result No.	Score	Query Match	Length	DB ID	Description
1	40	100.0	253	1 KLK7 HUMAN	P49862 homo sapien
2	34	85.0	201	1 FK11 MOUSE	Q9dim7 mus musculus
3	34	85.0	361	1 IHA TRIVU	Q77755 trichosaurus
4	33	82.5	339	1 YG47 XANAC	Q8p1y8 xanthomonas
5	33	82.5	1077	1 EM11 MOUSE	Q99k41 mus musculus
6	32	80.0	494	1 EXOT RHIME	P33699 rhizobium m
7	31	77.5	202	1 CUTI BOTCI	Q00298 botrytis ci
8	31	77.5	245	1 CG21 MOUSE	Q9jng3 mus musculus
9	31	77.5	246	1 CG21 HUMAN	Q9bvt8 homo sapien
10	31	77.5	354	1 GPDA TRYBB	P90593 trypanosoma
11	31	77.5	354	1 GPDA TRYBR	Q26756 trypanosoma
12	31	77.5	428	1 YFUD ECOLI	P37908 escherichia
13	31	77.5	783	1 RMHC DICDI	P34125 dictyosteli
14	30	75.0	155	1 IL2 BOVIN	P05016 bos taurus
15	30	75.0	155	1 IL2 BUBBU	Q95kp3 bubalus bub
16	30	75.0	155	1 IL2 CAPHI	P36835 capra hircu
17	30	75.0	157	1 IL2 SHREP	P19114 ovis aries
18	30	75.0	162	1 IL2 CEREL	P51747 cervus elap
19	30	75.0	289	1 CTR4 SCHFO	Q4722 schistosom
20	30	75.0	333	1 CD1B HUMAN	P29016 homo sapien
21	30	75.0	369	1 PSPD BOVIN	P35246 bos taurus
22	30	75.0	371	1 CL46 BOVIN	Q8mbz9 bos taurus
23	30	75.0	371	1 CONG BOVIN	P23805 bos taurus
24	30	75.0	374	1 YHUJ ECOLI	P31593 escherichia
25	30	75.0	435	1 IL49 HSVSA	Q01046 herpesvirus
26	30	75.0	433	1 VICE ECOLI	P27432 escherichia
27	30	75.0	524	1 K2C4 MOUSE	P07744 mus musculus
28	30	75.0	524	1 PPBT MOUSE	P09242 mus musculus
29	30	75.0	573	1 COX1 MYCTU	O53290 mycobacteri
30	30	75.0	592	1 S132 HUMAN	Q13183 homo sapien
31	30	75.0	783	1 RINI HUMAN	Q13671 homo sapien
32	30	75.0	902	1 NRE1 WOLSU	Q95le4 wolinnella s
33	30	75.0	1016	1 EM11 HUMAN	Q9y6c2 homo sapien

## ALIGNMENTS

### RESULT 1

ID	KLK7 HUMAN	STANDARD;	PRT;	253 AA.
AC	P49862;			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Kallikrein 7 precursor (EC 3.4.21.-) (Stratum corneum chymotryptic enzyme) (hSCCE).			
DE	enzyme (hSCCE).			
GN	KLK7 OR PRSS6 OR SCCE.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 23-53.			
RC	TISSUE=Skin;			
RX	MEDLINE=94308225; PubMed=8034709;			
RA	Hansson L., Stromqvist M., Baeckman A., Wallbrandt P., Carlstein A., Egelrud T.;			
RT	"Cloning, expression, and characterization of stratum corneum chymotryptic enzyme. A skin-specific human serine proteinase.";			
RL	J. Biol. Chem. 269:19420-19426(1994).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Keratinocytes;			
RA	Yousef G.M., Scorilas A., Diamandis E.P.;			
RT	"Molecular characterization, mapping and tissue expression of the human stratum corneum chymotryptic enzyme gene.";			
RL	Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=20510030; PubMed=11054574;			
RA	Gan L., Lee I., Smith R., Argonza-Barrett R., Lei H., McCuaig J., Moss P., Paepel B., Wang K.;			
RT	"Sequencing and expression analysis of the serine protease gene cluster located in chromosome 19q13 region.";			
RL	Gene 257:119-130(2000).			
RN	[4]			
RP	SEQUENCE FROM N.A.			
RA	Hansson L., Baeckman A., Ny A., Edlund M., Edholm E., Tornell J., Wallbrandt P., Egelrud T.;			
RT	"Epidermal overexpression of stratum corneum chymotryptic enzyme in mice; a model for chronic itchy dermatitis.";			
RL	Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.			
RN	[5]			
RP	CHARACTERIZATION.			
RX	MEDLINE=95314630; PubMed=7794273;			
RA	Skytt A., Stromqvist M., Egelrud T.;			
RT	"Primary substrate specificity of recombinant human stratum corneum chymotryptic enzyme.";			
RL	Biochem. Biophys. Res. Commun. 211:586-589(1995).			
CC	!- FUNCTION: May catalyze the degradation of intercellular cohesive structures in the cornified layer of the skin in the continuous shedding of cells from the skin surface. Specific for amino acid residues with aromatic side chains in the P1 position. SCCE			

P70553 rattus norv  
P30433 pongo pygma  
Q9pie4 campylobact  
O21515 phodopus su  
P01732 homo sapien  
P33706 canis famill  
Q44342 agrobacteri  
Q00825 odontella s  
P09326 homo sapien  
Q9jtx5 neisseria m  
Q91vvl neisseria m  
P81923 drosophila

34 29 72.5 126 1 NRM1 RAT  
35 29 72.5 198 1 CD8A\_PONPY  
36 29 72.5 202 1 Y357\_CAMJE  
37 29 72.5 208 1 NU4M\_PHOSU  
38 29 72.5 235 1 CD8A\_HUMAN  
39 29 72.5 239 1 CD8A\_CANFA  
40 29 72.5 239 1 FLGH\_AGRFS  
41 29 72.5 242 1 ATP1\_ODOSI  
42 29 72.5 243 1 CD48\_HUMAN  
43 29 72.5 306 1 TRUB\_NEIMA  
44 29 72.5 307 1 TRUB\_NEIMB  
45 29 72.5 397 1 O59A\_DROME

```

% cleaves insulin B chain at 6-Leu-|-Cys-7, 16-Tyr-|-Leu-17, 25-Phe-
% |Tyr-26, and 26-Tyr-|-Thr-27. Could play a role in the activation
% of precursors to inflammatory cytokines.
% -1- TISSUE SPECIFICITY: Abundantly expressed in the skin and is
% expressed by keratinocytes in the epidermis. Very low levels are
% also seen in the brain and kidney.
% -1- SIMILARITY: Belongs to peptidase family S1. Kallikrein subfamily.
%
% This SWISS-PROT entry is copyright. It is produced through a collaboration
% between the Swiss Institute of Bioinformatics and the EMBL outstation -
% the European Bioinformatics Institute. There are no restrictions on its
% use by non-profit institutions as long as its content is in no way
% modified and this statement is not removed. Usage by and for commercial
% entities requires a license agreement (See http://www.isb-sib.ch/announce/
% or send an email to license@isb-sib.ch).
%
% EMBL; L33404; AAC37551.1; --
% EMBL; AF166330; AAD49718.1; --
% EMBL; AF243527; AAG33360.1; --
% EMBL; AF232583; AAK69624.1; --
% PIR; A53968; A53968.
% HSP; P00763; IDPO.
% MEROPS; S01.300; -.
% Genes; HGNC:6368; KLK7.
% MIM; 604438; -.
% GO; GO:0008236; F-serine-type peptidase activity; TAS.
% GO; GO:0008544; P:epidermal differentiation; TAS.
% InterPro; IPR009003; Cys Ser trypsin.
% InterPro; IPR001254; Peptidase S1.
% InterPro; IPR001314; Peptidase_S1A.
% Pfam; PF00089; trypsin; 1.
% PRINTS; PR00722; CHYMOTRYPSIN.
% SMART; SM0020; TRYPSPC; 1.
% PROSITE; PS00240; TRYPSPIN_DOM; 1.
% PROSITE; PS00134; TRYPSPIN_HIS; 1.
% PROSITE; PS00135; TRYPSPIN_SRR; 1.
% Hydrolase; Serine protease; Zymogen; Glycoprotein; Signal.
% SIGNAL 1 22
% PROPEP 23 29 ACTIVATION PEPTIDE.
% CHAIN 30 253 KALLIKREIN 7.
% ACT_SITE 70 70 CHARGE RELAY SYSTEM (BY SIMILARITY).
% ACT_SITE 112 112 CHARGE RELAY SYSTEM (BY SIMILARITY).
% ACT_SITE 205 205 CHARGE RELAY SYSTEM (BY SIMILARITY).
% DISULFID 36 137 BY SIMILARITY.
% DISULFID 55 71 BY SIMILARITY.
% DISULFID 2 239 BY SIMILARITY.
% DISULFID 144 211 BY SIMILARITY.
% DISULFID 176 190 BY SIMILARITY.
% DISULFID 201 226 BY SIMILARITY.
% CARBOHYD 246 246 N-LINKED (GLCNAC...) (POTENTIAL).
% SEQUENCE 253 AA; 27525 MW; 2D69B6B15A76A568 CRC64;
%
% Query Match 100.0%; Score 40; DB 1; Length 253;
% Best Local Similarity 100.0%; Pred. No. 0.61; Indels 0; Gaps 0;
% Matches 9; Conservative 0; Mismatches 0;
%
% y 1 SLLPLQL 9
% 4 SLLPLQL 12
%
% RESULT 2
% FKIL_MOUSE STANDARD; PRT; 201 AA.
%
% C Q9D1N7; Q9CRE4;
% 28-FEB-2003 (Rel. 41, Created)
% 28-FEB-2003 (Rel. 41, Last sequence update)
% 15-MAR-2004 (Rel. 43, Last annotation update)
% FK506 binding protein 11 precursor (EC 5.2.1.8) (Peptidyl-prolyl cis-
% trans isomerase) (PPIase) (Rotamase) (19 kDa FK506-binding protein)
% (FKBP-19).
% N FKBP11.
% N Mus musculus (Mouse).

```

```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Embryo;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aitawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gotohori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schirrl L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barch G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohzuki S.,
RA Hayashizaki Y.;
RA "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Mammary gland;
RX MEDLINE=22389257; PubMed=1247932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong I.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Ioquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Robak S.A., Morley K.C., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs S.A.,
RA Fahey J., Helton E., Kettaman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -1- FUNCTION: PPIases accelerate the folding of proteins during
protein synthesis
CC -1- CATALYTIC ACTIVITY: Peptidylproline (omega=180) = peptidylproline
(omega=0).
CC -1- SIMILARITY: Belongs to the FKBP-type PPIase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AK003331; BAB22719.1; --
CC EMBL; AK019132; BAB31559.1; --
CC EMBL; BC037596; AAH37596.1; --
CC HSP; P20071; IYKJ
CC MGI; MGI:1933370; Fkbp11.
CC InterPro; IPR001179; FKBP_PPIase.
CC Pfam; PF00254; FKBP; 1.
CC PROSITE; PS00453; FKBP_PPIASE_1; 1.

```

DR PROSITE; PS00454; FKBP\_PPIASE\_2; FALSE\_NEG.  
 DR PROSITE; PS00059; FKBP\_PPIASE\_3; 1.  
 CW Isomerase; Rotamase; signal.  
 FT SIGNAL 1 27 POTENTIAL.  
 FT CHAIN 28 201 FK506 BINDING PROTEIN 11.  
 FT DOMAIN 57 144 PPIASE, FKBP-TYPE.  
 FT CONFLICT 53 53 S -> F (IN REF. 1; BAB31559).  
 FT CONFLICT 198 198 S -> R (IN REF. 1; BAB31559).  
 IQ SEQUENCE 201 AA; 22137 MW; 94D955C57264B082 CRC64;  
 Query Match 85.0%; Score 34; DB 1; Length 201;  
 Best Local Similarity 87.5%; Pred. No. 8.3; Indels 0; Gaps 0;  
 Matches 7; Conservative 1; Mismatches 0;  
 2 LLLPLQLL 9  
 6 LLLPLQLL 13  
 RESULT 3  
 LHA TRIVU STANDARD; PRT; 361 AA.  
 AC O7755;  
 JT 15-JUL-1999 (Rel. 38, Last sequence update)  
 JT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Inhibin alpha chain precursor.  
 INHA.  
 DS Trichosurus vulpecula (Brush-tailed possum).  
 SC Rukaryota; Metazoa; Chordata; Cranata; Vertebrata; Euteleostomi;  
 SC Mammalia; Metatheria; Diprotodontia; Phalangeridae; Trichosurus.  
 SC NCBI\_TaxID=9337;  
 SC SEQUENCE FROM N.A.  
 SC MEDLINE=99027340; PubMed=9801457;  
 SC Greenmont D., Fidler A.E., Heath D.A., Lawrence S.B., Tisdall D.J.,  
 SC "cDNA sequence analysis, gene expression and protein localisation of  
 SC the inhibin alpha subunit of Australian brushtail possum (Trichosurus  
 SC vulpecula).";  
 SC J. Mol. Endocrinol. 21:141-152(1998).  
 SC -1- FUNCTION: INHIBINS AND ACTIVINS INHIBIT AND ACTIVATE,  
 SC RESPECTIVELY, THE SECRETION OF FOLLITROPIN BY THE PITUITARY GLAND.  
 SC INHIBINS/ACTIVINS ARE INVOLVED IN REGULATING A NUMBER OF DIVERSE  
 SC FUNCTIONS SUCH AS HYPOTHALAMIC AND PITUITARY HORMONE SECRETION,  
 SC GONADAL HORMONE SECRETION, GERM CELL DEVELOPMENT AND MATURATION,  
 SC ERYTHROID DIFFERENTIATION, INSULIN SECRETION, NERVE CELL SURVIVAL,  
 SC EMBRYONIC AXIAL DEVELOPMENT OR BONE GROWTH, DEPENDING ON THEIR  
 SC SUBUNIT COMPOSITION. INHIBINS APPEAR TO OPPOSE THE FUNCTIONS OF  
 SC ACTIVINS.  
 SC -1- SUBUNIT: DIMERIC, LINKED BY ONE OR MORE DISULFIDE BONDS.  
 SC INHIBIN A IS A DIMER OF ALPHA AND BETA-A.  
 SC INHIBIN B IS A DIMER OF ALPHA AND BETA-B.  
 SC -1- SIMILARITY: Belongs to the TGF-beta family.  
 SC -----  
 SC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 SC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 SC the European Bioinformatics Institute. There are no restrictions on its  
 SC use by non-profit institutions as long as its content is in no way  
 SC modified and this statement is not removed. Usage by and for commercial  
 SC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 SC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 SC -----  
 SC ENBL; AF033340; AAC63945.1;  
 SC GO; GO:000576; C:extracellular; ISS.  
 SC GO; GO:001706; F:activin inhibitor activity; ISS.  
 SC GO; GO:0005125; F:cytokine activity; ISS.  
 SC GO; GO:0003793; F:defense/immunity protein activity; ISS.  
 SC GO; GO:0008083; F:growth factor activity; ISS.  
 SC GO; GO:0005179; F:hormone activity; ISS.  
 SC GO; GO:0005155; F:hormone binding; ISS.  
 SC GO; GO:0007050; P:cell cycle arrest; ISS.  
 SC GO; GO:0030154; P:cell differentiation; ISS.

DR GO; GO:0008151; P:cell growth and/or maintenance; ISS.  
 DR GO; GO:0007166; P:cell surface receptor linked signal transdu. .; ISS.  
 DR GO; GO:0007267; P:cell-cell signaling; ISS.  
 DR GO; GO:0030218; P:erythrocyte differentiation; ISS.  
 DR GO; GO:0042541; P:hemoglobin biosynthesis; ISS.  
 DR GO; GO:0006917; P:induction of apoptosis; ISS.  
 DR GO; GO:0045578; P:negative regulation of B-cell cell differentiation; ISS.  
 DR GO; GO:0045786; P:negative regulation of cell cycle; ISS.  
 DR GO; GO:0046892; P:negative regulation of follicle-stimulating. .; ISS.  
 DR GO; GO:0045077; P:negative regulation of interferon-gamma bio. .; ISS.  
 DR GO; GO:0045650; P:negative regulation of macrophage different. .; ISS.  
 DR GO; GO:0042328; P:negative regulation of phosphorylation; ISS.  
 DR GO; GO:0007399; P:neurogenesis; ISS.  
 DR GO; GO:0001541; P:ovarian follicle development; ISS.  
 DR GO; GO:0046881; P:positive regulation of follicle-stimulating. .; ISS.  
 DR InterPro; IPR002405; Inhibin\_alpha.  
 DR InterPro; IPR001839; TGFb.  
 DR Pfam; PF00019; TGF-beta; 1.  
 DR PRINTS; PR00669; INHIBIN.  
 DR PRODOM; PD000357; TGFb; 1.  
 DR SMART; SM00204; TGFb; 1.  
 DR PROSITE; PS00250; TGF\_BETA\_1; 1.  
 KW Growth factor; Hormone; Glycoprotein; signal.  
 FT SIGNAL 1 21 POTENTIAL.  
 FT PROPEP 22 230 BY SIMILARITY.  
 FT CHAIN 231 361 INHIBIN ALPHA CHAIN.  
 FT DISULFID 260 323 BY SIMILARITY.  
 FT DISULFID 289 358 BY SIMILARITY.  
 FT DISULFID 293 360 BY SIMILARITY.  
 FT DISULFID 322 322 INTERCHAIN (BY SIMILARITY).  
 FT CARBOHYD 48 48 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 144 144 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 266 266 N-LINKED (GLCNAC. .) (POTENTIAL).  
 SQ SEQUENCE 361 AA; 39345 MW; D661CDF93CDAA87D CRC64;  
 Query Match 85.0%; Score 34; DB 1; Length 361;  
 Best Local Similarity 87.5%; Pred. No. 15;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 2 LLLPLQLL 9  
 4 LLLPLQLL 11  
 Db  
 RESULT 4  
 YG47 XANAC STANDARD; PRT; 339 AA.  
 AC O8PLV8.  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Hypothetical protein XAC1647.  
 GN XAC1647.  
 OS Xanthomonas axonopodis (pv. citri).  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;  
 OC Xanthomonadaceae; Xanthomonas.  
 OC NCBI\_TaxID=92829;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=306 / ATCC 13902 / XV 101;  
 RX MEDLINE=2202145; PubMed=12024217;  
 RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,  
 RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,  
 RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,  
 RA Camarotte G., Cannavan F., Cardozo J., Chamberg F., Clapina L.P.,  
 RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,  
 RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,  
 RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,  
 RA Katsuvama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,  
 RA Locati E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,  
 RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,  
 RA Moreira L.M., Novo M.T.M., Okura V.R., Oliveira M.C., Oliveira V.R.,  
 RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,

A Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezra R.I.D.,  
A Trindade dos Santos M., Truffi D., Teai S.M., White F.F.,  
A Setubal J.C., Kitajima J.P.;  
T "Comparison of the genomes of two Xanthomonas pathogens with differing  
T host specificities";  
L Nature 417:459-463 (2002).  
C -!- SIMILARITY: Belongs to the UPF0190 family.  
C This SWISS-PROT entry is copyright. It is produced through a collaboration  
C between the Swiss Institute of Bioinformatics and the EMBL outstation -  
C the European Bioinformatics Institute. There are no restrictions on its  
C use by non-profit institutions as long as its content is in no way  
C modified and this statement is not removed. Usage by and for commercial  
C entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
C or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
R EMBL; AB011796; AAC36515.1; -;  
R HAMAP; MF 01206; -; 1.  
R InterPro; IPR000572; Oxidored\_mol.yb.  
R InterPro; IPR006311; Tat.  
R Pfam; PF00174; oxidored\_mol.yb; 1.  
R TIGRfam; TIGR01409; Tat\_signal\_seq; 1.  
W Hypothetical protein; Complete proteome.  
Q SEQUENCE 339 AA; 37901 MW; B926DFC9A8A84938 CRC64;  
Query Match 82.5%; Score 33; DB 1; Length 339;  
Best Local Similarity 88.9%; Pred No. 22;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Y 1 SLLPLQL 9  
b 5 SLLPLQL 13  
RESULT 5  
M1L MOUSE  
D EMBL\_MOUSE STANDARD; PRT; 1017 AA.  
C Q99K41;  
T 10-OCT-2003 (Rel. 42, Created)  
T 10-OCT-2003 (Rel. 42, Last sequence update)  
T 15-MAR-2004 (Rel. 43, Last annotation update)  
E EMBL1 precursor (Elastin microfibril interface-located protein 1)  
E (Elastin microfibril interface 1).  
N EMBL1.  
S Mus musculus (Mouse).  
C Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
C Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
X NCBI TaxID=10090;  
P [1]  
N P SEQUENCE FROM N.A.  
C STRAIN=C57BL/6J; TISSUE=Head;  
X MEDLINE=22354683; PubMed=12466951;  
A Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,  
A Nakaide I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,  
A Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,  
A Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,  
A Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,  
A Blake J.A., Bradt D., Brusci V., Chothia C., Corbani L.E., Cousins S.,  
A Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,  
A Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,  
A Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,  
A Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,  
A Kogagawa A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,  
A Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,  
A Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,  
A Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramchandran S.,  
A Ravasi T., Reed J.C., Reid J., Ring B.J., Ring B.Z., Ringwald M.,  
A Sandelin A., Schneider C., Sempile C.A., Setou M., Shimada K.,  
A Sultana R., Tanaka Y., Taylor M.S., Teasdale R.D., Tomita M.,  
A Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,  
A Wilming L.G., Wyshaw-Boris A., Yanagisawa M., Yang I., Yang L.,  
A Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,  
A Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,

RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,  
RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,  
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,  
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,  
RA Birney E., Hayashizaki Y.;  
RT "Analysis of the mouse transcriptome based on functional annotation of  
RT 60,770 full-length cDNAs";  
RL Nature 420:563-573 (2002).  
RN [2]  
RN SEQUENCE FROM N.A.  
RP TISSUE=Breast tumor;  
RC MEDLINE=22388257; PubMed=12477932;  
RX Srausberg R.L., Feigold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.W., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
RN [3]  
RN DEVELOPMENTAL STAGE.  
RX MEDLINE=22363881; PubMed=12475644;  
RA Bragheta P., Ferrari A., de Gemmis P., Zanetti M., Volpin D.,  
RA Bonaldo P., Bressan G.M.;  
RT "Expression of the EMLIN-1 gene during mouse development";  
RL Matrix Biol. 21:603-609 (2002).  
CC -!- FUNCTION: May be responsible for anchoring smooth muscle cells to  
CC elastic fibers, and may be involved not only in the formation of  
CC the elastic fiber, but also in the processes that regulate vessel  
CC assembly. Has cell adhesive capacity (By similarity). May have a  
CC function in placenta formation and initial organogenesis and a  
CC later role in interstitial connective tissue.  
CC -!- SUBUNIT: Homotrimer associated through a moderately stable  
CC interaction of the C-terminal globular C1q domains, allowing the  
CC nucleation of the triple helix and then a further quaternary  
CC assembly to higher order polymers via intermolecular disulfide  
CC bonds. Interacts with EMLIN2 (By similarity).  
CC -!- SUBCELLULAR LOCATION: Extracellular matrix. Found mainly at the  
CC interface between amorphous elastin and microfibrils.  
CC -!- DEVELOPMENTAL STAGE: Detectable in morula and blastocyst. First  
CC expressed in ectoplacental cone in embryos of 6.5 days and in  
CC extraembryonic visceral endoderm at 7.5 days. Expressed also in  
CC the allantois. Expression in the ectoplacental cone-derived  
CC secondary trophoblast giant cells and spongiotrophoblast is strong  
CC up to 11.5 days and then declines. In the embryo, high levels are  
CC initially expressed in blood vessels, perineural mesenchyme and  
CC somites at 8.5 days. Later on, intense expression is identified in  
CC the mesenchymal component of organs anlage (ie lung and liver) and  
CC different mesenchymal condensations (ie limb bud and branchial  
CC arches). At late gestation expression is widely distributed in  
CC interstitial connective tissue and smooth muscle cell-rich  
CC tissues.  
CC -!- SIMILARITY: Contains 1 C1q domain.  
CC -!- SIMILARITY: Contains 1 collagenous domain.  
CC -!- SIMILARITY: Contains 1 EMI domain.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way



modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; AK029337; BAC26403.1; --  
EMBL; BC005481; AAH05481.1; --  
MGD; MGI:1926189; Emilin1.  
InterPro; IPR001073; C1q.  
InterPro; IPR008160; Collagen.  
InterPro; IPR008983; TNF\_like.  
Pfam; PF00386; C1q; 1.  
Pfam; PF01391; Collagen; 1.  
Cell adhesion; Extracellular matrix; Signal; Coiled coil; Collagen;  
Glycoprotein. 1 23  
SIGNAL 24 1017  
CHAIN 24 1017  
DOMAIN 57 131  
DOMAIN 171 211  
DOMAIN 237 266  
DOMAIN 269 275  
DOMAIN 310 374  
DOMAIN 519 573  
DOMAIN 676 697  
DOMAIN 789 809  
DOMAIN 815 865  
DOMAIN 873 1008  
DOMAIN 156 156  
CARBOHYD 416 416  
CARBOHYD 456 456  
CARBOHYD 562 562  
CARBOHYD 659 659  
CARBOHYD 767 767  
CARBOHYD 795 795  
SEQUENCE 1017 AA; 107584 MW; 2EF903204DB9C88F CRC64;

Query Match 82.5%; Score 33; DB 1; Length 1017;  
Best Local Similarity 75.0%; Pred. No. 66;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Y 1 SLILPLQI 8  
b 973 SLILPLQV 980

RESULT 6  
XOT RHIME STANDARD; PRT; 494 AA.  
C P3369;  
T 01-FEB-1994 (Rel. 28, Created)  
T 28-FEB-2003 (Rel. 41, Last annotation update)  
E Succinoglycan biosynthesis transport protein exot.  
N EXOT OR RB1074 OR SM20950.  
S Rhizobium meliloti (Sinorhizobium meliloti).  
G Plasmid pSymb (megaplasmid 2).  
C Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
C Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.  
X NCBI\_TaxID=382;  
X [1]  
P SEQUENCE FROM N.A.  
C STRAIN=1021;  
X MEDLINE=94042869; PubMed=8226645;  
X Glucksmann M.A., Reuber T.L., Walker G.C.;  
T "Family of glycosyl transferases needed for the synthesis of  
T succinoglycan by Rhizobium meliloti.";  
T J. Bacteriol. 175:7033-7044 (1993).  
N [2]  
P SEQUENCE FROM N.A.  
C STRAIN=1021;  
X MEDLINE=94042870; PubMed=8226646;  
X Glucksmann M.A., Reuber T.L., Walker G.C.;  
T "Genes needed for the modification, polymerization, export, and  
T processing of succinoglycan by Rhizobium meliloti: a model for

RT succinoglycan biosynthesis.";  
RL J. Bacteriol. 175:7045-7055 (1993).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=RCR2011 / SU47;  
RX MEDLINE=94162682; PubMed=8118055;  
RA Becker A., Kleickmann A., Kuester H., Keller M., Arnold M.,  
RA Puchler A.;  
RT "Analysis of the Rhizobium meliloti genes exoU, exoV, exoW, exoT, and  
RT exoU involved in exopolysaccharide biosynthesis and nodule invasion:  
RT exoU and exoW probably encode glucosyltransferases.";  
RL Mol. Plant Microbe Interact. 6:735-744 (1993).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=1021;  
RX MEDLINE=21396508; PubMed=11481431;  
RA Finan T.M., Weidner S., Wong K., Buhrmester J., Chain P.,  
RA Vorhoelter F.J., Hernandez-Lucas I., Becker A., Gouzy J.,  
RA Golding B., Puchler A.;  
RT "The complete sequence of the 1.683-kb pSymb megaplasmid from the N2-  
RT fixing endosymbiont Sinorhizobium meliloti.";  
RL Proc. Natl. Acad. Sci. U.S.A. 98:9889-9894 (2001).  
CC -1- PATHWAY: Exopolysaccharide biosynthesis.  
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).  
CC -1- SIMILARITY: Belongs to the polysaccharide synthase family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; L20758; AA16050.1; --  
EMBL; Z22646; CA80362.1; --  
EMBL; AL503645; CAC49474.1; --  
DR PIR; B95976; B95976.  
DR PIR; C49349; C49349.  
DR InterPro; IPR002797; Polysacc\_synt.  
DR Pfam; PF01943; Polysacc\_synt; 1.  
KW Exopolysaccharide synthesis; Transmembrane; Plasmid;  
KW Complete proteome.  
FT TRANSMEM 16 36 POTENTIAL.  
FT TRANSMEM 44 64 POTENTIAL.  
FT TRANSMEM 82 102 POTENTIAL.  
FT TRANSMEM 105 125 POTENTIAL.  
FT TRANSMEM 157 177 POTENTIAL.  
FT TRANSMEM 215 235 POTENTIAL.  
FT TRANSMEM 253 273 POTENTIAL.  
FT TRANSMEM 297 317 POTENTIAL.  
FT TRANSMEM 321 341 POTENTIAL.  
FT TRANSMEM 343 363 POTENTIAL.  
FT TRANSMEM 384 404 POTENTIAL.  
FT TRANSMEM 421 441 POTENTIAL.  
FT TRANSMEM 447 467 POTENTIAL.  
SQ SEQUENCE 494 AA; 53315 MW; 008521D05FDB6013 CRC64;

Query Match 80.0%; Score 32; DB 1; Length 494;  
Best Local Similarity 75.0%; Pred. No. 52;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SLILPLQI 8  
Db 122 SLILPLQL 129

RESULT 7  
CUTI BOTCI STANDARD; PRT; 202 AA.  
ID CUTI BOTCI  
AC Q00298;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)

T 10-OCT-2003 (Rel. 42, Last annotation update)  
 N Cutinase precursor (SC 3.1.1.74) (Cutin hydrolase).  
 E CUTA.  
 S Botrytis cinerea (Noble rot fungus) (Botryotinia fuckeliana).  
 C Eukaryota; Fungi; Ascomycota; Pezizomycotina; Leotiomycetes;  
 X Helotiales; Sclerotiniaceae; Botryotinia.  
 C NCBI\_TaxID=40559;  
 N [1]  
 P SEQUENCE FROM N.A.  
 X STRAIN=SAS56;  
 X MEDLINE=97155570; PubMed=9002269;  
 A van der Vlugt-Bergmans C.J.B., Wagemakers L.C.A.M., van Kan J.A.L.;  
 T "Cloning and expression of the cutinase A gene of Botrytis cinerea.";  
 L Mol. Plant Microbe Interact. 10:21-29(1997).  
 C -!- FUNCTION: Catalyzes the hydrolysis of cutin, a polyester that  
 C forms the structure of plant cuticle. Allows pathogenic fungi to  
 C penetrate through the cuticular barrier into the host plant during  
 C the initial stage of the fungal infection (By similarity).  
 C -!- CATALYTIC ACTIVITY: Cutin + H(2)O = cutin monomers.  
 C -!- SUBCELLULAR LOCATION: Secreted (By similarity).  
 C -!- PTM: The 2 disulfide bonds play a critical role in holding the  
 C catalytic residues in juxta-position; reduction of the disulfide  
 C bridges results in the complete inactivation of the enzyme (By  
 C similarity).  
 C -!- SIMILARITY: Belongs to the cutinase family.  
 C  
 C This SWISS-PROT entry is copyright. It is produced through a collaboration  
 C between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 C the European Bioinformatics Institute. There are no restrictions on its  
 C use by non-profit institutions as long as its content is in no way  
 C modified and this statement is not removed. Usage by and for commercial  
 C entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 C or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 C  
 C EMBL; Z69264; CAA93255.1; -.  
 C HSP; P00590; 1XZJ.  
 C InterPro; IPR000675; Cutinase.  
 C Pfam; PF01083; Cutinase; 1.  
 C PRINTS; PR00129; CUTINASE.  
 C PROSITE; PS00155; CUTINASE\_1; 1.  
 C PROSITE; PS00931; CUTINASE\_2; 1.  
 C W Hydrolase; Serine esterase; Signal.  
 C T SIGNAL 1 20 POTENTIAL.  
 C T CHAIN 21 202 CUTINASE.  
 C T DISULFID 31 172 BY SIMILARITY.  
 C T DISULFID 106 165 BY SIMILARITY.  
 C T ACT SITE 117 117 BY SIMILARITY.  
 C T ACT SITE 169 169 BY SIMILARITY.  
 C T ACT SITE 182 182 BY SIMILARITY.  
 C Q SEQUENCE 202 AA; 20253 MW; 031D64725A23D5E7 CRC64;  
 Query Match 77.5%; Score 31; DB 1; Length 202;  
 Best Local Similarity 66.7%; Pred. No. 34;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 Y 1 SLILPLQL 9  
 b 11 ALLPLSLV 19  
 RESULT 8  
 D G21\_MOUSE STANDARD; PRT; 245 AA.  
 C Q9JWG3;  
 T 10-OCT-2003 (Rel. 42, Created)  
 T 10-OCT-2003 (Rel. 42, Last sequence update)  
 T 10-OCT-2003 (Rel. 42, Last annotation update)  
 E Protein C7orf21 homolog.  
 N Mus musculus (Mouse).  
 S Mus musculus.  
 C Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 C Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 X NCBI\_TaxID=10090;  
 RN SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=20145471; PubMed=10679242;  
 RA Inoue S., Sano H., Ohta M.;  
 RT "Growth suppression of Escherichia coli by induction of expression of  
 RT mammalian genes with transmembrane or ATPase domains.";  
 RL Biochem. Biophys. Res. Commun. 268:553-561(2000).  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Small intestine;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itch M., Ishii Y.,  
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 RA Azawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,  
 RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadoya K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
 RA Schram L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Moshima J., Mazzarelli J., Momota N.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Suzuki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kotsuki S.,  
 RA Hayashizaki Y.;  
 RT "Functional annotation of a full-length mouse cDNA collection.";  
 RL Nature 409:685-690(2001).  
 RN SEQUENCE FROM N.A.  
 RC TISSUE=Salivary gland;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
 RA Schnerch A., Schein J.E., Jones S.U.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
 CC -!- SIMILARITY: Contains 1 ubiquitin-like domain.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AB030183; BAA92747.1; -.  
 DR EMBL; AK008109; BAB25465.1; -.  
 DR EMBL; BC019547; AAR19547.1; -.  
 DR MGD; MGI:1923764; 2010004O20Rik.  
 DR InterPro; IPR000626; Ubiquitin.  
 DR Pfam; PF00240; ubiquitin; 1.

```

R SMART: SM00213; UBO; 1.
R PROSITE; PS00553; UBIQUITIN_2; 1.
W Transmembrane.
T T TRANSMEM 11 31 POTENTIAL.
T T TRANSMEM 194 214 POTENTIAL.
T T TRANSMEM 219 239 POTENTIAL.
T T TRANSMEM 102 175 UBIQUITIN-LIKE.
T T DOMAIN 197 238 LEU-RICH.
T T DOMAIN 245 AA; 26316 MW; 02CID1B24C13139D CRC64;
Q SEQUENCE 245 AA; 26316 MW; 02CID1B24C13139D CRC64;

Query Match 77.5%; Score 31; DB 1; Length 245;
Best Local Similarity 77.8%; Pred. No. 42;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Y 1 SLLPLQIL 9
b 196 SLLPLLLL 204

RESULT 9
G21_HUMAN
D G21_HUMAN STANDARD; PRT; 246 AA.
C C Q9V78;
T 10-OCT-2003 (Rel. 42, Created)
T 10-OCT-2003 (Rel. 42, Last sequence update)
T 10-OCT-2003 (Rel. 42, Last annotation update)
E Protein C7orf21 (SB144).
N C7ORF21.
S Homo sapiens (Human).
C Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
C Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
X NCBI_TaxID=9606;
N N
P SEQUENCE FROM N.A.
L Zhang W., Li N., Wan T., Zhang M., Cao X.;
L Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
P [2]
P SEQUENCE FROM N.A.
C TISSUE=Kidney, and Placenta;
X MEDLINE=22398257; PubMed=12477932;
X Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
X Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
X Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
X Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
X Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
X Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
X Brownstein M.J., Uedin T.B., Toshiyuki S., Carninci P., Prange C.,
X Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
X Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
X Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
X Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
X Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
X Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
X Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
X Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
X Butterfield Y.S.N., Krzywicki M.I., Skalska U., Smallos D.E.,
X Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
T "Generation and initial analysis of more than 15,000 full-length human
T and mouse cDNA sequences.";
L Proc. Natl. Acad. Sci. U.S.A. 98:15689-15693(2002).
C -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
C -1- SIMILARITY: Contains 1 ubiquitin-like domain.
C -----
C This SWISS-PROT entry is copyright. It is produced through a collaboration
C between the Swiss Institute of Bioinformatics and the EMBL outstation -
C the European Bioinformatics Institute. There are no restrictions on its
C use by non-profit institutions as long as its content is in no way
C modified and this statement is not removed. Usage by and for commercial
C entities requires a license agreement (See http://www.isb-sib.ch/announce/
C or send an email to license@isb-sib.ch).
C -----
R EMBL; AY037155; AAK67645.1; -.
R EMBL; BC000936; AAH00936.1; -.

```

```

DR EMBL; BC033182; AAH33182.1; -.
DR Genew; HGNC:21709; C7orf21.
DR InterPro; IPR000626; Ubiquitin.
DR Pfam; PF00240; ubiquitin; 1.
DR PROSITE; PS00553; UBIQUITIN_2; 1.
FT Transmembrane.
FT TRANSMEM 11 31 POTENTIAL.
FT TRANSMEM 195 215 POTENTIAL.
FT TRANSMEM 221 241 POTENTIAL.
FT DOMAIN 103 176 UBIQUITIN-LIKE.
FT DOMAIN 198 238 LEU-RICH.
SQ SEQUENCE 246 AA; 26261 MW; E08E25AGB37665B3 CRC64;

Query Match 77.5%; Score 31; DB 1; Length 246;
Best Local Similarity 77.8%; Pred. No. 42;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 SLLPLQIL 9
Db 197 SLLPLLLL 205

RESULT 10
GPDA_TRYBB
ID GPDA_TRYBB STANDARD; PRT; 354 AA.
AC P05953;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Glycerol-3-phosphate dehydrogenase [NAD+], glycosomal (EC 1.1.1.8).
GN GPD.
OS Trypanosoma brucei brucei.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
CX NCBI_TaxID=5702;
RN SEQUENCE FROM N.A.
RP STRAIN=427;
RX MEDLINE=97077437; PubMed=8920004;
RA Kohl L., Dmota T., Thi C.D., Callens M., van Beeumen J.,
RA Oppenoes F.R., Michels P.A.M.;
RT "Cloning and characterization of the NAD-linked glycerol-3-phosphate
RT dehydrogenases of Trypanosoma brucei and Leishmania mexicana
RT mexicana and expression of the trypanosome enzyme in Escherichia
RT coli.";
RL Mol. Biochem. Parasitol. 76:159-173(1996).
CC -1- CATALYTIC ACTIVITY: Sn-glycerol 3-phosphate + NAD(+) = glycerone
CC phosphate + NADH.
CC -1- SUBCELLULAR LOCATION: Glycosomal.
CC -1- SIMILARITY: Belongs to the NAD-dependent glycerol-3-phosphate
CC dehydrogenase family.
C -----
C This SWISS-PROT entry is copyright. It is produced through a collaboration
C between the Swiss Institute of Bioinformatics and the EMBL outstation -
C the European Bioinformatics Institute. There are no restrictions on its
C use by non-profit institutions as long as its content is in no way
C modified and this statement is not removed. Usage by and for commercial
C entities requires a license agreement (See http://www.isb-sib.ch/announce/
C or send an email to license@isb-sib.ch).
C -----
DR EMBL; X89738; CAA61890.1; -.
DR PIR; T48649; T48649.
DR InterPro; IPR008927; EDGDH C-like.
DR InterPro; IPR006168; NAD_Gly3P_dh.
DR Pfam; PF01210; NAD_Gly3P_dh; 1.
DR PRINTS; PR00077; GPDHGNASE.
DR PRODOM; PD001278; NAD_Gly3P_dh; 1.
DR PROSITE; PS00957; NAD_G3PDH; 1.
DR Oxidoreductase; NAD; Glycosome.
KW SITE 352 354 MICROBODY TARGETING SIGNAL (POTENTIAL).
FT SEQUENCE 354 AA; 37783 MW; 12CE19A5DA9E4EC9 CRC64;

Query Match 77.5%; Score 31; DB 1; Length 354;

```

```

Best Local Similarity 77.8%; Pred. No. 60;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Y 1 SLLPLQIL 9
b 124 SLLFPAQIL 132

RESULT 11
PDA TRYBR
C Q26756; STANDARD; PRT; 354 AA.
T 30-MAY-2000 (Rel. 39, Created)
T 30-MAY-2000 (Rel. 39, Last sequence update)
T 15-MAY-2004 (Rel. 43, Last annotation update)
E Glycero1-3-phosphate dehydrogenase [NAD+], glycosomal (EC 1.1.1.8).
N GPD.
S Trypanosoma brucei rhodesiense.
C Eukaryota; Eucleozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
X NCBI_TaxID=31286;
X [1]
P SEQUENCE FROM N.A.
C STRAIN=Vitat 1.1;
X MEDLINE=97077436; PubMed=8920003;
A Stebeck C.B., Frevet U., Mommsen T.P., Vassella E., Roditi I.,
P Pearson T.W.;
T "Molecular characterization of glycosomal NAD(+) dependent glycerol 3-
T phosphate dehydrogenase from Trypanosoma brucei rhodesiense."
L Mol. Biochem. Parasitol. 76:145-158(1996).
C -!- CATALYTIC ACTIVITY: Sn-glycerol 3-phosphate + NAD(+) = glycerone
C phosphate + NADH.
C -!- SUBCELLULAR LOCATION: Glycosomal.
C -!- SIMILARITY: Belongs to the NAD-dependent glycerol-3-phosphate
C dehydrogenase family.
C
C This SWISS-PROT entry is copyright. It is produced through a collaboration
C between the Swiss Institute of Bioinformatics and the EMBL outstation -
C the European Bioinformatics Institute. There are no restrictions on its
C use by non-profit institutions as long as its content is in no way
C modified and this statement is not removed. Usage by and for commercial
C entities requires a license agreement (See http://www.isb-sib.ch/announce/
C or send an email to license@isb-sib.ch)
C
C EMBL; X91142; CA62581.1; -
C InterPro; IPR008927; 6DGDH_C like.
C InterPro; IPR006168; NAD Gly3P dh.
C InterPro; IPR006109; NAD Gly3P dh.
C Pfam; PF01210; NAD Gly3P dh; 1.
C PRINTS; PR00077; GPDHGNASE.
C ProDom; PD001278; NAD Gly3P dh; 1.
C PROSITE; PS00957; NAD_G3PDR; 1.
W Oxidoreductase; NAD; Glycosome.
I SITE 352 354 MICROBODY TARGETING SIGNAL (POTENTIAL).
Q SEQUENCE 354 AA; 37834 MW; 0CB049C5E8F03C00 CRC64;

Query Match 77.5%; Score 31; DB 1; Length 354;
Best Local Similarity 77.8%; Pred. No. 60;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Y 1 SLLPLQIL 9
b 124 SLLFPAQIL 132

RESULT 12
FUD ECOLI
C YFUD_ECOLI STANDARD; PRT; 428 AA.
C P37908; P76600; P76601; P77009;
T 01-OCT-1994 (Rel. 30, Created)
T 15-DEC-1998 (Rel. 37, Last sequence update)
T 16-OCT-2001 (Rel. 40, Last annotation update)
E Hypothetical protein yfjd.
F Complete proteome.
Y YFUD OR B2612/B2613.

```

```

OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474 (1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97349980; PubMed=9205837;
RA Yamamoto Y., Aiba H., Baba T., Hayashi K., Inada T., Isono K.,
RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,
RA Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nashimoto H.,
RA Oshima T., Oyama S., Saito N., Sampei G., Satoh Y., Sivasubram S.,
RA Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,
RA Yamagata S., Horiuchi T.;
RT "Construction of a contiguous 874-kb sequence of the Escherichia coli
RT - K12 genome corresponding to 50.0-68.8 min on the linkage map and
RT analysis of its sequence features."
RL DNA Res. 4:91-113(1997).
RN [3]
RP SEQUENCE OF 322-428 FROM N.A.
RC STRAIN=B178;
RX MEDLINE=88319942; PubMed=3045760;
RA Lipinska B., King J., Aug D., Georgopoulos C.;
RT "Sequence analysis and transcriptional regulation of the Escherichia
RT coli grpE gene, encoding a heat shock protein."
RL Nucleic Acids Res. 16:7545-7562(1988).
RN [4]
RP IDENTIFICATION.
RC Rudd K.E.;
RX Unpublished observations (AUG-1994).
RL -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: BELONGS TO THE UPF0053 FAMILY. STRONG, TO
CC H.INFLUENZAE H10107.
CC -!- SIMILARITY: Contains 2 CBS domains.
CC -!- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A
CC FRAMESHIFT IN POSITION 199 THAT PRODUCES TWO ORFS.
CC -!- CAUTION: Ref.3 sequence differs from that shown due to numerous
CC frameshifts.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC
CC EMBL; A5000347; AAC75661.1; ALT_FRAME.
CC EMBL; A5000347; AAC75662.1; ALT_FRAME.
CC EMBL; D30888; BAAL6497.1; ALT_INIT.
CC EMBL; X07863; -; NOT ANNOTATED_CDS.
CC EcoGene; EGI2442; yfjd.
CC InterPro; IPR002550; CBS.
CC InterPro; IPR000644; CBS domain.
CC InterPro; IPR005170; CorC_transpt-asc.
CC Pfam; PF00571; CBS; 2.
CC Pfam; PF03471; CorC_HlyC; 1.
CC Pfam; PF01595; DUF21; 1.
CC SMART; SM00116; CBS; 1.
CC Hypothetical protein; CBS domain; Repeat; Transmembrane;
KW Complete proteome.
FT TRANSMEM 4 24 POTENTIAL.
FT 65 85 POTENTIAL.

```

T TRANSMEM 92 112 POTENTIAL.  
T TRANSMEM 130 150 POTENTIAL.  
Q SEQUENCE 428 AA; 48044 MW; 3845685EDBCB151 CRC64;  
Query Match 77.5%; Score 31; DB 1; Length 428;  
Best Local Similarity 87.5%; Pred. No. 72;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Y 2 LLLPQL 9  
b 131 LLLPQL 138  
RESULT 13  
KHC\_DICDI STANDARD; PRT; 783 AA.  
D KHC\_DICDI  
C F34125;  
T 01-FEB-1994 (Rel. 28, Created)  
T 01-OCT-1994 (Rel. 30, Last sequence update)  
T 15-MAR-2004 (Rel. 43, Last annotation update)  
T MYOSIN heavy chain kinase (EC 2.7.1.129) (MHCK).  
E MHCK.  
N Dictyostelium discoideum (Slime mold).  
X Dictyostelium discoideum (Slime mold).  
X Eukaryota; Mycetozoa; Dictyostelidae; Dictyostelium.  
X NCBI\_TaxID=44689;  
U SEQUENCE FROM N.A.  
U STRAIN=AX3;  
U MEDLINE=92335200; PubMed=1321427;  
U Ravid S., Spudich J.A.;  
U "Membrane-bound Dictyostelium myosin heavy chain kinase: a developmentally regulated substrate-specific member of the protein kinase C family.";  
U Proc. Natl. Acad. Sci. U.S.A. 89:5877-5881(1992).  
U -!- FUNCTION: Brings about phosphorylation of specific threonines in the heavy chains of myosin during chemotaxis, which results in inhibition of myosin thick filament formation. Phosphorylation also inhibits the actin-activated ATPase activity of myosin.  
U -!- CATALYTIC ACTIVITY: ATP + [myosin heavy-chain] = ADP + [myosin heavy-chain] phosphate.  
U -!- SUBCELLULAR LOCATION: Membrane-bound.  
U -!- DEVELOPMENTAL STAGE: Expressed only during development.  
U -!- SIMILARITY: Belongs to the eukaryotic diacylglycerol kinase family.  
U -!- SIMILARITY: Contains 2 zinc-dependent phorbol-ester and DAG binding domains.  
U -!- CAUTION: While this is experimentally defined to be a protein kinase, it could be, on the basis of sequence similarities, a diacylglycerol kinase.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
-----  
EMBL; M93393; -; NOT\_ANNOTATED\_CDS.  
PIR; A46136; A46136.  
HSSP; P04049; 1FAR.  
DictyBase; DDB0185000; M93393.  
InterPro; IPR002219; DAG\_PE-bind.  
InterPro; IPR000756; DAGKa.  
InterPro; IPR001206; DAGKc.  
Pfam; PF00130; DAG\_PE-bind; 1.  
Pfam; PF00603; DAGKa; 1.  
Pfam; PF00781; DAGKc; 1.  
ProDom; PD002939; DAGKa; 1.  
ProDom; PD005043; DAGKc; 1.  
SMART; SM00109; C1; 2.  
SMART; SM00045; DAGKa; 1.  
SMART; SM00046; DAGKc; 1.

DR PROSITE; PS00479; DAG\_PE\_BIND\_DOM 1; 1.  
DR PROSITE; PS50081; DAG\_PE\_BIND\_DOM 2; 2.  
KW Transferase; Serine/threonine-protein kinase; ATP-binding; Membrane; phosphorylation; Phorbol-ester binding; Repeat.  
KW DOMAIN 19 69 PHORBOL-ESTER AND DAG BINDING 1.  
FT NP\_BIND 88 138 PHORBOL-ESTER AND DAG BINDING 2.  
FT BINDING 466 474 ATP (BY SIMILARITY).  
FT BINDING 489 489 ATP (BY SIMILARITY).  
FT DOMAIN 188 196 ASN-RICH.  
FT MOD\_RES 754 773 SER/THR-RICH.  
FT MOD\_RES 768 768 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
SQ SEQUENCE 783 AA; 89154 MW; 6CAF177F4436F43E CRC64;  
Query Match 77.5%; Score 31; DB 1; Length 783;  
Best Local Similarity 77.8%; Pred. No. 1.3e+02;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 1 SLLLPQL 9  
Db 236 SLLLPQL 244  
RESULT 14  
IL2\_BOVIN STANDARD; PRT; 155 AA.  
ID IL2\_BOVIN  
AC P05016;  
DT 13-AUG-1987 (Rel. 05, Created)  
DT 13-AUG-1987 (Rel. 05, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Interleukin-2 precursor (IL-2) (T-cell growth factor) (TCGF).  
GN IL2 OR IL-2.  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.  
OX NCBI\_TaxID=9913;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=86205869; PubMed=3517854;  
RA Cerretti D.P., McKereghan K., Larsen A., Cantrell M.A., Anderson D., Gillis S., Cosman D., Baker P.E.;  
RA "Cloning, sequence, and expression of bovine interleukin 2.";  
RT Proc. Natl. Acad. Sci. U.S.A. 83:3223-3227(1986).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=86205870; PubMed=3486415;  
RA Reeves R., Spies A.G., Nissen M.S., Buck C.D., Weinberg A.D., Barr P.J., Magnuson N.S., Magnuson J.A.;  
RA "Molecular cloning of a functional bovine interleukin 2 cDNA.";  
RT Proc. Natl. Acad. Sci. U.S.A. 83:3228-3232(1986).  
RN [3]  
RP SEQUENCE OF 1-22 FROM N.A.  
TISSUE=Thymus;  
RC Anikeeva N.N., Vincogradova T.V., Votoshin O.N.;  
RL Submitted (DEC-1989) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: Produced by T-cells in response to antigenic or mitogenic stimulation, this protein is required for T-cell proliferation and other activities crucial to regulation of the immune response. Can stimulate B cells, monocytes, lymphokine-activated killer cells, natural killer cells, and glioma cells.  
CC -!- SUBCELLULAR LOCATION: Secreted.  
CC -!- SIMILARITY: Belongs to the IL-2 family.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
-----  
EMBL; M12791; AAA30586.1; -; ALT\_INIT.  
EMBL; M13204; AAA21143.1; -;

```
3 EMBL; X17201; CAA35062.1; -.
4 EMBL; X52687; CAA36912.1; -.
5 PIR; I45913; I45913.
6 HSP; P01585; 3INK.
7 InterPro; IPR000779; Interleukin-2.
8 Pfam; PF00715; IL2; 1.
9 PRINTS; PR00265; INTERLEUKIN2.
10 ProDom; PD003649; Interleukin-2; 1.
11 SMART; SM00189; IL2; 1.
12 PROSITE; PS00424; INTERLEUKIN_2; 1.
13 Cytokine; Glycoprotein; Immune response; Signal; Growth factor;
14 T-cell.
15 SIGNAL 1 20 BY SIMILARITY.
16 CHAIN 21 155 INTERLEUKIN-2.
17 DISULFID 79 127 BY SIMILARITY.
18 CARBOHYD 23 23 O-LINKED (GALNAC. .) (BY SIMILARITY).
19 CONFLICT 66 66 V -> A (IN REF. 2).
20 SEQUENCE 155 AA; 17627 MW; 816667DFA052EDF CRC64;

Query Match 75.0%; Score 30; DB 1; Length 155;
Best Local Similarity 77.8%; Pred. No. 43;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Y 1 SLLPLQL 9
b 36 SLLDLQL 44
```

Search completed: March 1, 2004, 17:29:59  
Job time : 7 secs

```
Y 1 SLLPLQL 9
b 36 SLLDLQL 44
```

```
RESULT 15
L2_BUBBU STANDARD; PRT; 155 AA.
C Q95XP3;
T 10-OCT-2003 (Rel. 42, Created)
T 10-OCT-2003 (Rel. 42, Last sequence update)
T 10-OCT-2003 (Rel. 42, Last annotation update)
E Interleukin-2 precursor (IL-2) (T-cell growth factor) (TCGF).
N IL2
S Bubalus bubalis (Domestic water buffalo).
C Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
C Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
C Bovidae; Bovinae; Bubalus.
X NCBI_TaxID=89462;
X [1]
P SEQUENCE FROM N.A.
X MEDLINE=22117321; PubMed=12121282;
A Sreekumar E., Premraj A., Saravankumar M., Rasool T.J.;
T "Buffalo (Bubalus bubalis) interleukin-2: sequence analysis reveals
T high nucleotide and amino acid identity with interleukin-2 of cattle
T and other ruminants."
L Eur. J. Immunogenet. 29:341-345(2002).
C -!- FUNCTION: Produced by T-cells in response to antigenic or
C mitogenic stimulation, this protein is required for T-cell
C proliferation and other activities crucial to regulation of the
C immune response. Can stimulate B cells, monocytes, lymphokine-
C activated killer cells, natural killer cells, and glioma cells (By
C similarity).
C -!- SUBCELLULAR LOCATION: Secreted.
C -!- SIMILARITY: Belongs to the IL-2 family.
C -----
C This SWISS-PROT entry is copyright. It is produced through a collaboration
C between the Swiss Institute of Bioinformatics and the EMBL outstation -
C the European Bioinformatics Institute. There are no restrictions on its
C use by non-profit institutions as long as its content is in no way
C modified and this statement is not removed. Usage by and for commercial
C entities requires a license agreement (See http://www.isb-sib.ch/announce/
C or send an email to license@isb-sib.ch).
C -----
C EMBL; AF363786; AAK50039.1; -.
R InterPro; IPR000779; Interleukin-2.
R Pfam; PF00715; IL2; 1.
R PRINTS; PR00265; INTERLEUKIN2.
R ProDom; PD003649; Interleukin-2; 1.
R SMART; SM00189; IL2; 1.
```

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

M protein - protein search, using sw model

run on: March 1, 2004, 17:20:41 ; Search time 30.3333 Seconds  
(without alignments)  
93.615 Million cell updates/sec

file: US-09-905-083-36  
effect score: 40  
sequence: 1 SLLPLQL 9

coring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

searched: 1017041 seqs, 315518202 residues

total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

database :

1: SPREMBL.25.\*

2: sp\_archaea.\*

3: sp\_bacteria.\*

4: sp\_fungi.\*

5: sp\_human.\*

6: sp\_invertebrate.\*

7: sp\_mammal.\*

8: sp\_mhc.\*

9: sp\_organelle.\*

10: sp\_phase.\*

11: sp\_plant.\*

12: sp\_rodent.\*

13: sp\_virus.\*

14: sp Vertebrate.\*

15: sp\_unclassified.\*

16: sp\_virus.\*

17: sp\_bacteriap.\*

18: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	40	100.0	253	4 Q8NSN9	Q8NSN9 homo sapien
2	38	95.0	146	17 Q9V2D5	Q9V2D5 pyrococcus
3	37	92.5	208	8 Q21527	Q21527 clethrionom
4	34	85.0	73	11 Q8VCA9	Q8VCA9 mus musculus
5	34	85.0	104	11 Q8R5D6	Q8R5D6 mus musculus
6	33	82.5	370	16 Q8YHY9	Q8YHY9 bruceella me
7	33	82.5	370	16 Q8FZX1	Q8FZX1 bruceella su
8	33	82.5	379	16 Q7U9J6	Q7U9J6 synchococcc
9	33	82.5	381	16 Q7V4U4	Q7V4U4 prochloroco
10	33	82.5	352	16 Q819M8	Q819M8 bacillus ce
11	32	80.0	72	12 Q8VBB6	Q8VBB6 white spot
12	32	80.0	78	2 Q7WXY8	Q7WXY8 pseudomonas
13	32	80.0	102	11 Q8BRV0	Q8BRV0 mus musculus
14	32	80.0	171	16 Q7U5R4	Q7U5R4 synchococcc
15	32	80.0	196	11 Q9D226	Q9D226 mus musculus
16	32	80.0	218	10 Q9ZTE3	Q9ZTE3 arabidopsis

Q876R3 emericeella  
Q922G5 mus musculus  
Q8UIE2 agrobacteri  
Q9E8S3 mus musculus  
Q354Z6 mus musculus  
Q9R1S4 rattus norv  
Q9M0J5 arabidopsis  
Q861E6 porphyromon  
Q821L4 streptomyc  
Q12212 saccharomyc  
Q969D0 giardia lam  
Q8X56 arabidopsis  
Q9F149 arabidopsis  
Q8LWX9 laminaria d  
Q9P1I9 homo sapien  
Q8PKF6 xanthomonas  
Q8ZUD7 pyrobaculum  
Q9VYN0 drosophila  
Q9X620 salmonella  
Q761S5 periplaneta  
Q84BF5 xanthomonas  
Q8WX77 homo sapien  
Q22910 arabidopsis  
Q9V3V4 drosophila  
Q8MMA8 syringa sib  
Q8UEY8 agrobacteri  
Q82VH0 nitrosomona  
Q9KAL1 bacillus ha  
P96675 bacillus su

#### ALIGNMENTS

#### RESULT 1

Q8NSN9 PRELIMINARY; PRT; 253 AA.  
AC Q8NSN9;  
DT 01-OCT-2002 (TRENBLrel. 22, Created)  
DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)  
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)  
DE Kallikrein 7 (chymotryptic, stratum corneum).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Skin;  
RA Strausberg R.;  
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.  
DR EMBL; BC032005; AAH32005.1; -;  
DR GO; GO:0004263; F:chymotrypsin activity; IEA.  
DR GO; GO:0008233; F:peptidase activity; IEA.  
DR GO; GO:0004295; F:trypsin activity; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR InterPro; IPR009003; Cys Ser trypsin.  
DR InterPro; IPR001254; Peptidase S1.  
DR InterPro; IPR001314; Peptidase\_S1A.  
DR Pfam; PF00089; trypsin; 1.  
DR PRINTS; PR00722; CHYMOTRYPSIN.  
DR SMART; SMO0020; TRY\_SPC; 1.  
DR PROSITE; PS0240; TRYPsin\_DOM; 1.  
DR PROSITE; PS00134; TRYPsin\_HIS; 1.  
DR PROSITE; PS00135; TRYPsin\_SER; 1.  
DR Hydrolase; Protease; Serine protease.  
KW SEQUENCE 253 AA; 27608 MW; 2D68B6A41B2A668 CRC64;

Query Match 100.0%; Score 40; DB 4; Length 253;  
Best Local Similarity 100.0%; Pred. No. 3.7;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Y 1 SLLPLQIL 9
b 4 SLLPLQIL 12

RESULT 2
Q9V2D5 PRELIMINARY; PRT; 146 AA.
D Q9V2D5;
T 01-MAY-2000 (TrEMBLrel. 13, Created)
T 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
T 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
E Hypothetical protein PYRAB01390.
N PYRAB01390 OR PAB0088.
S Pyrococcus abyssi.
C Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
C Pyrococcus.
X NCBI_TaxID=29292;
[1]
N SEQUENCE FROM N.A.
P STRAIN=GE5 / Orsay;
A Heilig R.;
T "Pyrococcus abyssi genome sequence: insights into archaeal chromosome
structure and evolution.";
L Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
R EMBL; AJ248283; CAB49063.1; -.
R F1R; H75201; H75201.
R InterPro; IPR008537; DUF819.
R Pfam; PF05684; DUF819; 1.
W Hypothetical protein; Complete proteome.
Q SEQUENCE 146 AA; 16092 MW; 718294137125801F CRC64;

Query Match 95.0%; Score 38; DB 17; Length 146;
Best Local Similarity 88.9%; Pred. No. 5.7;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Y 1 SLLPLQIL 9
b 66 SLLPLQII 74

RESULT 3
Q21527 PRELIMINARY; PRT; 208 AA.
D Q21527;
T 01-JAN-1998 (TrEMBLrel. 05, Created)
T 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
T 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
E NADH dehydrogenase subunit 4 (EC 1.6.5.3) (NADH-ubiquinone
oxidoreductase chain 4) (Fragment).
N ND4.
S Clethrionomys gapperi (Southern red-backed vole).
X Mitochondrion
X Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
X Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Arvicolinae;
X Clethrionomys.
X NCBI_TaxID=56223;
[1]
N SEQUENCE FROM N.A.
P MEDLINE=98152303; PubMed=9491603;
A Engel S.R., Hogan K.M., Taylor J.F., Davis S.K.;
T "Molecular systematics and paleobiogeography of the South American
sigmodontine rodents.";
L Mol. Biol. Evol. 15:35-49(1998).
C -/- CATALYTIC ACTIVITY: NADH + UBIQUINONE = NAD(+) + UBIQUINOL.
R EMBL; U83808; AAB87168.1; -.
R GO; GO:0005739; F:Mitochondrion; IEA.
R GO; GO:0008137; F:NADH dehydrogenase (ubiquinone) activity; IEA.
R GO; GO:0016491; F:oxidoreductase activity; IEA.
R GO; GO:0006120; P:mitochondrial electron transport, NADH to u. . .; IEA.
R InterPro; IPR003918; NADH_oxred4.
R InterPro; IPR001750; Oxidored_q1.
R InterPro; IPR000260; Oxidored_q5_N.

DR Pfam; PF00361; oxidored_q1; 1.
DR Pfam; PF01059; oxidored_q5_N; 1.
DR PRINTS; PR01437; NCOXDRDTASE4.
KW NAD; Oxidoreductase; Ubiquinone; Mitochondrion.
FT NON TER 208
SQ SEQUENCE 208 AA; 23967 MW; 8AF1788697AED6A2 CRC64;

Query Match 92.5%; Score 37; DB 8; Length 208;
Best Local Similarity 88.9%; Pred. No. 13;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SLLPLQIL 9
Db 97 SLLVPLQIL 105

RESULT 4
Q8VCA9 PRELIMINARY; PRT; 73 AA.
ID Q8VCA9
AC Q8VCA9;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Similar to RIKEN cDNA 1110002023 gene.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
[1]
N SEQUENCE FROM N.A.
P TISSUE=Salivary gland;
R Strausberg R.;
L Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
R EMBL; BC021345; AAB21345.1; -.
SQ SEQUENCE 73 AA; 7819 MW; 9358F53399BF3C11 CRC64;

Query Match 85.0%; Score 34; DB 11; Length 73;
Best Local Similarity 87.5%; Pred. No. 20;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LLLPLQIL 9
Db 6 LLLPLQLL 13

RESULT 5
Q8RSD6 PRELIMINARY; PRT; 104 AA.
ID Q8RSD6
AC Q8RSD6;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Similar to RIKEN cDNA 1110002023 gene.
GN FKBP11 OR 1110002023RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
[1]
N SEQUENCE FROM N.A.
P Strausberg R.;
R Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
R EMBL; BC022900; AAB22900.1; -.
DR MGD; MGI:1913370; FKBP11.
DR GO; GO:0006457; P:protein folding; IEA.
DR InterPro; IPR001179; FKBP_PPIase.
DR Pfam; PF00254; FKBP; 1.
DR PROSITE; PS00453; FKBP_PPIASE_1; 1.
DR PROSITE; PS00059; FKBP_PPIASE_3; 1.
SQ SEQUENCE 104 AA; 11085 MW; 0534D57467566914 CRC64;

Query Match 85.0%; Score 34; DB 11; Length 104;
Best Local Similarity 87.5%; Pred. No. 27;

```



Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Y 2 LLLPLQL 9  
 :|||||:  
 6 LLLPLQL 13

RESULT 6  
 QYHY9 PRELIMINARY; PRT; 370 AA.

ID QYHY9  
 AC QYHY9; 01-MAR-2002 (TREMBLrel. 20, Created)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)  
 DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)  
 DE Daunorubicin resistance transmembrane protein.  
 EN BMEI0656  
 OS Brucella melitensis.  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 OC Brucellaceae; Brucella.  
 OC NCBI\_TaxID=29459;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=16M / ATCC 23456 / Biotype 1;  
 RX MEDLINE=20020103; PubMed=175668;  
 SA Delvecchio V.G., Kapatal V., Redkar R.J., Patra G., Mujar C., Los T.,  
 SA Ivanova N., Anderson I., Bhattacharya A., Lykidis A., Reznik G.,  
 SA Jablonski L., Larsen N., D'Souza M., Bernal A., Mazur M., Goltzman E.,  
 SA Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Letesson J.-J.,  
 SA Haselkorn R., Kyrides N., Overbeek R.;  
 SA "The genome sequence of the facultative intracellular pathogen  
 SA Brucella melitensis"; U.S.A. 99:443-448(2002).  
 RT Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).  
 DR EMBL; AE009507; AAU51837.1; -;  
 DR PIR; AB3334; AB3334.  
 GW Complete proteome.  
 SQ SEQUENCE 370 AA; 40369 MW; 819AAD694305CB60 CRC64;

Query Match 82.5%; Score 33; DB 16; Length 370;  
 Best Local Similarity 75.0%; Pred. No. 1.4e+02;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Y 2 LLLPLQL 9  
 :|||||:  
 293 LLLPLQL 300

RESULT 7  
 QYHY9 PRELIMINARY; PRT; 370 AA.

ID QYHY9  
 AC QYHY9; 01-MAR-2002 (TREMBLrel. 23, Created)  
 DT 01-MAR-2002 (TREMBLrel. 23, Last sequence update)  
 DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)  
 DE ABC transporter, permease protein, putative.  
 EN BR1349.  
 OS Brucella suis.  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 OC Brucellaceae; Brucella.  
 OC NCBI\_TaxID=29461;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=1330 / Biovar 1;  
 RX MEDLINE=22247741; PubMed=122711122;  
 SA Paulsen I.T., Seshadri R., Unayam L., Brinkac L.M., Beanan M.J.,  
 SA Read T.D., Dodson R.J., Durkin A.S., Kolonay J.F., Madupu R.,  
 SA Daugherty S.C., Deboy R.T., Kaul M., Shetty J., Malek J., Van Aken S.E.,  
 SA Nelson W.C., Ayodeji B., Kral M., Shetty J., Malek J., Van Aken S.E.,  
 SA Riedmuller S., Rietlin H., Gill S.R., White O., Salzberg S.L.,  
 SA Hoover D.L., Lindler L.E., Hailing S.M., Boyle S.M., Fraser C.M.;  
 SA "The Brucella suis genome reveals fundamental similarities between  
 SA animal and plant pathogens and symbionts."  
 RT Proc. Natl. Acad. Sci. U.S.A. 99:13148-13153(2002).  
 RL EMBL; AE014432; AAN30263.1; -;

DR PIR; AB3334; AB3334.  
 DR TIGR; BR1349; -;  
 KW Complete proteome.  
 SQ SEQUENCE 370 AA; 40369 MW; 819AAD694305CB60 CRC64;

Query Match 82.5%; Score 33; DB 16; Length 370;  
 Best Local Similarity 75.0%; Pred. No. 1.4e+02;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LLLPLQL 9  
 :|||||:  
 293 LLLPLQL 300

RESULT 8  
 QYHY9 PRELIMINARY; PRT; 379 AA.

ID QYHY9  
 AC QYHY9; 01-OCT-2003 (TREMBLrel. 25, Created)  
 DT 01-OCT-2003 (TREMBLrel. 25, Last sequence update)  
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
 DE Putative glycosyltransferase precursor (EC 2.4.1.-).  
 EN SYN00260.  
 OS Synechococcus sp. (strain WH8102).  
 OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.  
 OC NCBI\_TaxID=84588;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22825697; PubMed=12917641;  
 RA Palenik B., Brahmsha B., Larimer F.W., Land M., Hauser L., Chain P.,  
 RA Lamerdin J., Regala W., Allen E.E., McCarrren J., Paulsen I.,  
 RA Dufresne A., Partensky F., Webb E.A., Waterbury J.;  
 RA "The genome of a motile marine Synechococcus";  
 RT Nature 424:1037-1042(2003).  
 RL EMBL; BX569689; CAE06775.1; -;  
 DR EMBL; BX569689; CAE06775.1; -;  
 FT SIGNAL 19  
 SQ SEQUENCE 379 AA; 40174 MW; 98D697CECF71C73F CRC64;

Query Match 82.5%; Score 33; DB 16; Length 379;  
 Best Local Similarity 88.9%; Pred. No. 1.5e+02;  
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 SLLPLQL 9  
 :|||||:  
 257 SLLPLQL 265

RESULT 9  
 QYHY9 PRELIMINARY; PRT; 381 AA.

ID QYHY9  
 AC QYHY9; 01-OCT-2003 (TREMBLrel. 25, Created)  
 DT 01-OCT-2003 (TREMBLrel. 25, Last sequence update)  
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
 DE Glycosyl transferase, family 4 precursor (EC 2.4.1.-).  
 EN PRT1846.  
 OS Prochlorococcus marinus (strain MIT 9313).  
 OC Bacteria; Cyanobacteria; Prochlorophytes; Prochlorococcaceae;  
 OC Prochlorococcus.  
 OC NCBI\_TaxID=74547;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22825698; PubMed=12917642;  
 RA Rocap G., Larimer F.W., Lamerdin J., Malfatti S., Chain P.,  
 RA Ahlgren N.A., Arriano A., Coleman M., Hauser L., Hess W.R.,  
 RA Johnson Z.I., Land M., Lindell D., Post A.F., Regala W., Shah M.,  
 RA Shaw S.L., Steglich C., Sullivan M.B., Ting C.S., Tolonen A.,  
 RA Webb E.A., Zinser E.R., Chisholm S.W.;  
 RA "Genome divergence in two Prochlorococcus ecotypes reflects oceanic  
 RA niche differentiation."  
 RT Nature 424:1042-1047(2003).  
 RL EMBL; BX572100; CAE22021.1; -;

M Glycosyltransferase; Signal; Transferase; Complete proteome.  
T SIGNAL 1 19  
Q SEQUENCE 381 AA; 40217 MW; 4DBAB55F313F8BE2 CRC64;

Query Match 82.5%; Score 33; DB 16; Length 381;  
Best Local Similarity 88.9%; Pred. No. 1.5e+02;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

NY 1 SLLPLQL 9  
b 257 SLLPLLIL 265

RESULT 10

819M8 Q819M8 PRELIMINARY; PRT; 392 AA.  
C Q819M8  
D 01-JUN-2003 (TrEMBLrel. 24, Created)  
E 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
F 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
G Cell division protein ftsW.  
H BC3948.  
I Bacillus cereus (strain ATCC 14579 / DSM 31).  
J Bacteria; Firmicutes; Bacilliales; Bacillaceae; Bacillus.  
K NCBI\_TaxID=226300;  
L [1]  
M SEQUENCE FROM N.A.  
N MEDLINE=22608415; PubMed=12721630;  
O Ivanova N., Sorokin A., Anderson I., Galleron N., Candelon B.,  
P Kapural V., Bhattacharyya A., Reznik G., Mikhailova N., Lapidus A.,  
Q Chu L., Mazur M., Goltzman E., Larsen N., D'Souza M., Walunas T.,  
R Grechkin Y., Pusch G., Haselkorn R., Fonstein M., Ehrlich S.D.,  
S Overbeek R., Kyrpides N.  
T "Genome sequence of *Bacillus cereus* and comparative analysis with  
U *Bacillus anthracis*."  
V Nature 423:87-91(2003).  
W EMBL; AE017011; AAP10868.1; --  
X GO; GO:0016020; C:membrane; IEA.  
Y GO; GO:0005311; F:sodiumdicarboxylate/tricarboxylate symport. . .; IEA.  
Z GO; GO:0000910; P:cytokinesis; IEA.  
aa GO; GO:0006835; P:dicarboxylic acid transport; IEA.  
ab InterPro; IPR001182; C:cell cycle.  
ac InterPro; IPR001991; Na/diCO symport.  
ad Pfam; PF01098; FTSW RODA SPOVE; 1.  
ae PRINTS; PR00173; EDTNSPORT.  
af PROSITE; PS00428; FTSW RODA SPOVE; 1.  
ag PROSITE; PS00012; PHOSPHOPANTHEINE; 1.  
ah Cell division; Complete proteome.  
ai SEQUENCE 392 AA; 42797 MW; 7AA558737762F659 CRC64;

Query Match 82.5%; Score 33; DB 16; Length 392;  
Best Local Similarity 88.9%; Pred. No. 1.5e+02;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

NY 1 SLLPLQL 9  
b 11 SLLPLVIL 19

RESULT 11

8VB86 Q8VB86 PRELIMINARY; PRT; 72 AA.  
C Q8VB86  
D 01-MAR-2002 (TrEMBLrel. 20, Created)  
E 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
F 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
G Wav052 (WSSV109).  
H White spot syndrome virus (WSSV).  
I Viruses; dsDNA viruses, no RNA stage; Nimaviridae.  
J NCBI\_TaxID=92652;  
K [1]  
L SEQUENCE FROM N.A.

RX MEDLINE=21548311; PubMed=11889662;  
RA Yang F., He J., Lin X., Li Q., Pan D., Zhang X., Xu X.;  
R "Complete genome sequence of the shrimp white spot bacilliform  
RT virus."  
RN J. Virol. 75:11811-11820(2001).  
RL (2)  
RM SEQUENCE FROM N.A.  
RA Yang F., He J., Lin X., Li Q., Pan D., Zhang X., Xu X.;  
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.  
RN (3)  
RM SEQUENCE FROM N.A.  
RA STRAIN=Taiwan;  
RC MEDLINE=20517548; PubMed=11062040;  
RX Tsai M.F., Yu H.T., Tseng H.F., Leu J.H., Chou C.M., Huang C.J.,  
RA Wang C.H., Lin J.Y., Kou G.H., Lo C.F.;  
R "Identification and characterization of a shrimp white spot syndrome  
RT virus (WSSV) gene that encodes a novel chimeric polypeptide of  
cellular-type thymidine kinase and thymidylate kinase."  
RL Virology 277:100-110(2000).  
RN (4)  
RM SEQUENCE FROM N.A.  
RA STRAIN=Taiwan;  
RC MEDLINE=21844071; PubMed=11853398;  
RX Chen L.L., Leu J.H., Huang C.J., Chou C.M., Chen S.M., Wang C.H.,  
RA Lo C.F., Kou G.H.;  
R "Identification of a nucleocapsid protein (VP35) gene of shrimp white  
RT spot syndrome virus and characterization of the motif important for  
targeting VP35 to the nuclei of transfected insect cells."  
RL Virology 293:44-53(2002).  
RN (5)  
RM SEQUENCE FROM N.A.  
RA STRAIN=Taiwan;  
RC Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
RL EMBL; AF332093; AAL33056.1; --  
DR EMBL; AF40570; AAL8897.1; --  
SQ SEQUENCE 72 AA; 7862 MW; E84A9814F2D7581A CRC64;

Query Match 80.0%; Score 32; DB 12; Length 72;  
Best Local Similarity 100.0%; Pred. No. 50;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 LLLPLQI 8  
Db 35 LLLPLQI 41

RESULT 12

Q7WXY8 Q7WXY8 PRELIMINARY; PRT; 78 AA.  
AC Q7WXY8;  
DT 01-OCT-2003 (TrEMBLrel. 25, Created)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Hypothetical protein.  
CN PL072  
OS Pseudomonas aeruginosa.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
OC Pseudomonadaceae; Pseudomonas.  
OX NCBI\_TaxID=287;  
RN [1]  
RM SEQUENCE FROM N.A.  
RC STRAIN=PA14;  
RA He J., Rahme L.G.;  
R "Pseudomonas aeruginosa PA14 pathogenicity island (PAPI) 1."  
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY273869; AAP84197.1; --  
KW Hypothetical protein.  
SQ SEQUENCE 78 AA; 8977 MW; 9B37EF214B4C91A1 CRC64;

Query Match 80.0%; Score 32; DB 2; Length 78;  
Best Local Similarity 77.8%; Pred. No. 54;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

M protein - protein search, using sw model

un on: March 1, 2004, 17:21:46 ; Search time 11.8899 Seconds  
(without alignments)  
39.081 Million cell updates/sec

title: US-09-905-083-80

effect score: 50

sequence: 1 GPLVCRGTL 9

coring table: BLOSJMC2

Gapop 10.0 , Gapext 0.5

searched: 389414 seqs, 51625971 residues

total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A COMB.pep:\*  
2: /cgn2\_6/ptodata/2/iaa/5B COMB.pep:\*  
3: /cgn2\_6/ptodata/2/iaa/6A COMB.pep:\*  
4: /cgn2\_6/ptodata/2/iaa/6B COMB.pep:\*  
5: /cgn2\_6/ptodata/2/iaa/PTUS COMB.pep:\*  
6: /cgn2\_6/ptodata/2/iaa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

result No.	Score	Query Match	Length	DB ID	Description
1	50	100.0	9	3	US-09-502-600-80
2	50	100.0	9	4	US-09-918-243-80
3	50	100.0	224	3	US-08-944-483-33
4	50	100.0	225	2	US-08-557-146-12
5	50	100.0	225	2	US-09-027-337-4
6	50	100.0	225	2	US-09-154-344-12
7	50	100.0	225	4	US-09-644-600-4
8	50	100.0	225	4	US-09-654-600A-4
9	50	100.0	253	2	US-08-557-146-2
10	50	100.0	253	2	US-08-824-874-3
11	50	100.0	253	2	US-09-154-344-2
12	50	100.0	253	3	US-08-930-188-2
13	50	100.0	253	3	US-09-210-084-3
14	50	100.0	253	3	US-09-764-762-3
15	50	100.0	253	5	PCT-US96-04294-2
16	41	82.0	3201	4	US-09-679-379-15
17	41	82.0	3546	4	US-09-679-379-13
18	40	80.0	20	1	US-08-472-228A-19
19	40	80.0	20	4	US-09-146-831-19
20	40	80.0	20	5	PCT-US96-09303-19
21	40	80.0	237	1	US-08-096-946-10
22	40	80.0	237	1	US-08-096-946-11
23	40	80.0	237	2	US-08-844-024-2
24	40	80.0	237	2	US-08-718-547-2
25	40	80.0	237	3	US-08-768-859A-1
26	40	80.0	237	3	US-08-768-859A-16
27	40	80.0	237	3	US-08-768-859A-21

Sequence 1, Appli  
Sequence 16, Appli  
Sequence 21, Appli  
Sequence 1, Appli  
Sequence 7, Appli  
Sequence 12, Appli  
Sequence 37, Appli  
Sequence 38, Appli  
Sequence 1, Appli  
Sequence 3, Appli  
Sequence 12, Appli  
Sequence 2, Appli  
Sequence 1, Appli  
Sequence 7, Appli  
Sequence 8, Appli  
Sequence 1, Appli  
Sequence 10, Appli  
Sequence 11, Appli

28 40 80.0 237 3 US-08-767-820A-1  
29 40 80.0 237 3 US-08-767-820A-16  
30 40 80.0 237 3 US-08-767-820A-21  
31 40 80.0 237 3 US-08-622-046B-1  
32 40 80.0 237 3 US-08-622-046B-7  
33 40 80.0 237 3 US-08-622-046B-12  
34 40 80.0 237 3 US-08-944-483-37  
35 40 80.0 237 3 US-08-944-483-38  
36 40 80.0 237 3 US-09-100-264-1  
37 40 80.0 237 3 US-09-100-264-3  
38 40 80.0 237 3 US-09-100-264-12  
39 40 80.0 237 4 US-09-303-339-2  
40 80.0 237 4 US-08-843-076D-1  
41 40 80.0 237 4 US-08-843-076D-7  
42 40 80.0 237 4 US-08-843-076D-8  
43 40 80.0 237 4 US-09-303-208-1  
44 40 80.0 237 5 PCT-US94-07329-10  
45 40 80.0 237 5 PCT-US94-07329-11

#### ALIGNMENTS

RESULT 1  
US-09-502-600-80  
; Sequence 80, Application US/09502600A  
; Patent No. 6294344

GENERAL INFORMATION:

APPLICANT: O'Brien, Timothy J.

TITLE OF INVENTION: Compositions and Methods for the Early Diagnosis of

TITLE OF INVENTION: Ovarian Cancer

FILE REFERENCE: D6223CIP-C

CURRENT FILING DATE: 2000-02-11

CURRENT APPLICATION NUMBER: US/09/502,600A

PRIOR APPLICATION NUMBER: 09/039,211

PRIOR FILING DATE: 03-14-1998

NUMBER OF SEQ ID NOS: 136

SEQ ID NO 80

LENGTH: 9

TYPE: PRT

ORGANISM: Homo sapiens

FEATURE:

OTHER INFORMATION: Residues 207-215 of the SCCE protein

US-09-502-600-80

Query Match 100.0%; Score 50; DB 3; Length 9;

Best Local Similarity 100.0%; Pred. No. 3e+05;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GPLVCRGTL 9

DB 1 GPLVCRGTL 9

RESULT 2

US-09-918-243-80

; Sequence 80, Application US/09918243

; Patent No. 6627403

GENERAL INFORMATION:

APPLICANT: O'Brien, Timothy J.

APPLICANT: Cannon, Martin J.

APPLICANT: Santin, Alessandro

TITLE OF INVENTION: Methods for the early diagnosis of ovarian cancer

FILE REFERENCE: D6223CIP/C/D/CIP

CURRENT APPLICATION NUMBER: US/09/918,243

CURRENT FILING DATE: 2001-07-30

PRIOR APPLICATION NUMBER: US

PRIOR FILING DATE: 2001-07-13

NUMBER OF SEQ ID NOS: 136

SEQ ID NO 80

LENGTH: 9

TYPE: PRT

ORGANISM: Homo sapiens

## FEATURE:

NAME/KEY: CHAIN

OTHER INFORMATION: Residues 207-215 of the SCCE protein

S-09-918-243-80

Query Match 100.0%; Score 50; DB 4; Length 9;  
Best Local Similarity 100.0%; Pred. No. 3e+05;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

y 1 GPLVCRGTL 9

b 1 GPLVCRGTL 9

## RESULT 3

S-08-944-483-33

Sequence 33, Application US/08944483

Patent No. 6232456

GENERAL INFORMATION:

APPLICANT: COHEN, MAURICE

APPLICANT: COLPITTS, TRACEY L.

APPLICANT: FRIEDMAN, PAULA N.

APPLICANT: GRANADOS, EDWARD N.

APPLICANT: KLASS, MICHAEL R.

APPLICANT: RUSSELL, JOHN C.

APPLICANT: STEWART, KENT D.

APPLICANT: STROUPE, STEVEN D.

TITLE OF INVENTION: NOVEL SERINE PROTEASE REAGENTS

TITLE OF INVENTION: AND METHODS USEFUL FOR DETECTING AND TREATING DISEASES

NUMBER OF SEQUENCES: 76

CORRESPONDENCE ADDRESS:

ADDRESS: Abbott Laboratories

STREET: 100 Abbott Park Road

CITY: Abbott Park

STATE: IL

COUNTRY: USA

ZIP: 60064-3500

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/944,483

FILING DATE:

CLASSIFICATION: 424

PRIOR APPLICATION NUMBER:

APPLICATION NUMBER:

FILING DATE:

NAME: Becker, Cheryl L.

REGISTRATION NUMBER: 35,441

REFERENCE/DOCKET NUMBER: 6183.US.01

TELEPHONE: 847/935-1729

TELEFAX: 847/938-2623

TELEX:

INFORMATION FOR SEQ ID NO: 33:

SEQUENCE CHARACTERISTICS:

LENGTH: 224 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: No. 6232456e

S-08-944-483-33

Query Match 100.0%; Score 50; DB 3; Length 224;  
Best Local Similarity 100.0%; Pred. No. 0.14;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

y 1 GPLVCRGTL 9

b 1 GPLVCRGTL 9

Db 178 GPLVCRGTL 186

## RESULT 4

US-08-557-146-12

Sequence 12, Application US/08557146

Patent No. 5834290

GENERAL INFORMATION:

APPLICANT: Egelrud, Torbjorn

APPLICANT: Hansson, Lennart

TITLE OF INVENTION: Recombinant Stratum Corneum Chymotryptic

TITLE OF INVENTION: Enzyme (SCCE)

NUMBER OF SEQUENCES: 17

CORRESPONDENCE ADDRESS:

ADDRESSEE: White &amp; Case, Patent Department

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: U.S.A.

ZIP: 10036-2787

COMPUTER READABLE FORM: disk

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.25 (EPO)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/557,146

FILING DATE: 14-DEC-1995

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: Sterner, Richard J.

REGISTRATION NUMBER: 35,372

REFERENCE/DOCKET NUMBER: 1103326-181

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 819-8783

TELEFAX: (212) 354-8113

INFORMATION FOR SEQ ID NO: 12:

SEQUENCE CHARACTERISTICS:

LENGTH: 225 amino acids

TYPE: amino acids

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: polypeptide

US-08-557-146-12

Query Match 100.0%; Score 50; DB 2; Length 225;  
Best Local Similarity 100.0%; Pred. No. 0.14;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GPLVCRGTL 9

Db 179 GPLVCRGTL 187

Qy 1 GPLVCRGTL 9

Db 179 GPLVCRGTL 187

Qy 1 GPLVCRGTL 9

Db 179 GPLVCRGTL 187

Qy 1 GPLVCRGTL 9

Db 179 GPLVCRGTL 187

Qy 1 GPLVCRGTL 9

Db 179 GPLVCRGTL 187

Qy 1 GPLVCRGTL 9

Db 179 GPLVCRGTL 187

Qy 1 GPLVCRGTL 9

Db 179 GPLVCRGTL 187

Qy 1 GPLVCRGTL 9

Db 179 GPLVCRGTL 187

Qy 1 GPLVCRGTL 9

Db 179 GPLVCRGTL 187

Qy 1 GPLVCRGTL 9

Db 179 GPLVCRGTL 187

Qy 1 GPLVCRGTL 9

Db 179 GPLVCRGTL 187

Qy 1 GPLVCRGTL 9

Db 179 GPLVCRGTL 187

Qy 1 GPLVCRGTL 9

Db 179 GPLVCRGTL 187

Qy 1 GPLVCRGTL 9

Db 179 GPLVCRGTL 187

Qy 1 GPLVCRGTL 9

Db 179 GPLVCRGTL 187

Qy 1 GPLVCRGTL 9

Db 179 GPLVCRGTL 187

Qy 1 GPLVCRGTL 9

Db 179 GPLVCRGTL 187

OTHER INFORMATION: similar domain in TADG-15  
JS-09-027-337-4

Query Match 100.0%; Score 50; DB 2; Length 225;  
Best Local Similarity 100.0%; Pred. No. 0.14; 0; Indels 0; Gaps 0;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2y 1 GPLVCRGTL 9  
179 GPLVCRGTL 187

## RESULT 6

JS-09-154-344-12  
Sequence 12, Application US/09154344  
Patent No. 5981256  
GENERAL INFORMATION:  
APPLICANT: Egelrud, Torbjorn  
APPLICANT: Hansson, Lennart  
TITLE OF INVENTION: Recombinant Stratum Corneum Chymotryptic  
TITLE OF INVENTION: Enzyme (SCCE)  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: White & Case, Patent Department  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2787  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25 (BPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/154,344  
FILING DATE: 16-SEP-1998  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/557,146  
FILING DATE: 14-DEC-1995  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Sterner, Richard J.  
REGISTRATION NUMBER: 35,372  
REFERENCE/DOCKET NUMBER: 1103326-181  
TELEPHONE: (212) 819-8783  
TELEFAX: (212) 354-8113  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 225 amino acids  
TYPE: amino acids  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: polypeptide  
JS-09-154-344-12

Query Match 100.0%; Score 50; DB 2; Length 225;  
Best Local Similarity 100.0%; Pred. No. 0.14; 0; Indels 0; Gaps 0;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2y 1 GPLVCRGTL 9  
179 GPLVCRGTL 187

## RESULT 7

JS-09-644-600-4  
Sequence 4, Application US/09644600  
Patent No. 6451500  
GENERAL INFORMATION:  
APPLICANT: O'Brien, Timothy J.

APPLICANT: Tanimoto, Hirotooshi  
TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease  
FILE REFERENCE: D6064CIP/D  
CURRENT APPLICATION NUMBER: US/09/644,600  
CURRENT FILING DATE: 2000-08-23  
PRIOR APPLICATION NUMBER: 09/421,213  
PRIOR FILING DATE: 1999-10-20  
PRIOR APPLICATION NUMBER: 09/027,337  
PRIOR FILING DATE: 1998-02-20  
NUMBER OF SEQ ID NOS: 98  
SEQ ID NO 4  
LENGTH: 225  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: SCCE  
US-09-644-600-4

Query Match 100.0%; Score 50; DB 4; Length 225;  
Best Local Similarity 100.0%; Pred. No. 0.14; 0; Indels 0; Gaps 0;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GPLVCRGTL 9  
Db 179 GPLVCRGTL 187

## RESULT 8

US-09-654-600A-4  
Sequence 4, Application US/09654600A  
Patent No. 6649741  
GENERAL INFORMATION:  
APPLICANT: O'Brien, Timothy J.  
APPLICANT: Tanimoto, Hirotooshi  
TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease  
FILE REFERENCE: D6064CIP/D  
CURRENT APPLICATION NUMBER: US/09/654,600A  
CURRENT FILING DATE: 2000-09-01  
PRIOR APPLICATION NUMBER: 09/421,213  
PRIOR FILING DATE: 1999-10-20  
NUMBER OF SEQ ID NOS: 98  
SEQ ID NO 4  
LENGTH: 225  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: SCCE  
US-09-654-600A-4

Query Match 100.0%; Score 50; DB 4; Length 225;  
Best Local Similarity 100.0%; Pred. No. 0.14; 0; Indels 0; Gaps 0;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GPLVCRGTL 9  
Db 179 GPLVCRGTL 187

## RESULT 9

US-08-557-146-2  
Sequence 2, Application US/08557146  
Patent No. 5834290  
GENERAL INFORMATION:  
APPLICANT: Egelrud, Torbjorn  
APPLICANT: Hansson, Lennart  
TITLE OF INVENTION: Recombinant Stratum Corneum Chymotryptic  
TITLE OF INVENTION: Enzyme (SCCE)  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:

ADDRESSEE: White & Case, Patent Department  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2787  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/557,146  
FILING DATE: 14-DEC-1995  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Steiner, Richard J.  
REGISTRATION NUMBER: 35,372  
REFERENCE/DOCKET NUMBER: 1103326-181  
TELEPHONE: (212) 819-8783  
TELEFAX: (212) 354-8113  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 253 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-557-146-2

Query Match 100.0%; Score 50; DB 2; Length 253;  
Best Local Similarity 100.0%; Pred. No. 0.16;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 GPLVCRGTL 9  
| | | | |  
Db 207 GPLVCRGTL 215

RESULT 10  
US-08-824-874-3  
Sequence 3, Application US/08824874  
Patent No. 5962300  
GENERAL INFORMATION:  
APPLICANT: Hillman, Jennifer L.  
APPLICANT: Lal, Preeti  
TITLE OF INVENTION: NOVEL KALLIKREIN  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/824,874  
FILING DATE: Filed Herewith  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0252 US  
TELEPHONE: 415-855-0555

TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 253 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 532504  
US-08-824-874-3

Query Match 100.0%; Score 50; DB 2; Length 253;  
Best Local Similarity 100.0%; Pred. No. 0.16;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 GPLVCRGTL 9  
| | | | |  
Db 207 GPLVCRGTL 215

RESULT 11  
US-09-154-344-2  
Sequence 2, Application US/09154344  
Patent No. 5981256  
GENERAL INFORMATION:  
APPLICANT: Egelrud, Torbjorn  
APPLICANT: Hasson, Lennart  
TITLE OF INVENTION: Recombinant Stratum Corneum Chymotryptic  
TITLE OF INVENTION: Enzyme (SCCE)  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: White & Case, Patent Department  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2787  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/154,344  
FILING DATE: 16-SEP-1998  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/557,146  
FILING DATE: 14-DEC-1995  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Sterner, Richard J.  
REGISTRATION NUMBER: 35,372  
REFERENCE/DOCKET NUMBER: 1103326-181  
TELEPHONE: (212) 819-8783  
TELEFAX: (212) 354-8113  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 253 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-154-344-2

Query Match 100.0%; Score 50; DB 2; Length 253;  
Best Local Similarity 100.0%; Pred. No. 0.16;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 GPLVCRGTL 9  
| | | | |  
Db 207 GPLVCRGTL 215

## RESULT 12

S-08-930-188-2

Sequence 2, Application US/089301088

Patent No. 6093397

GENERAL INFORMATION:

APPLICANT: Dixon, Eric P.

APPLICANT: Johnstone, Edward M.

APPLICANT: Little, Shella P.

TITLE OF INVENTION: AMYLOID PRECURSOR PROTEIN PROTEASE AND

TITLE OF INVENTION: RELATED NUCLEIC ACIDS

NUMBER OF SEQUENCES: 3

CORRESPONDENCE ADDRESS:

ADDRESSER: Eli Lilly and Company

STREET: Lilly Corporate Center

CITY: Indianapolis

STATE: Indiana

COUNTRY: United States of America

ZIP: 46285

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/930,188

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/416,257

FILING DATE: 04-APR-1995

ATTORNEY/AGENT INFORMATION:

NAME: Blalock, Donna K.

REGISTRATION NUMBER: 38,082

REFERENCE/DOCKET NUMBER: X9239

TELECOMMUNICATION INFORMATION:

TELEPHONE: 317-277-1090

TELEFAX: 317-276-3861

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 253 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

JS-08-930-188-2

Query Match 100.0%; Score 50; DB 3; Length 253;

Best Local Similarity 100.0%; Pred. No. 0.16; Indels 0; Gaps 0;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY

1 GPLVCRGTL 9

|||||

207 GPLVCRGTL 215

DB

## RESULT 13

US-09-210-084-3

Sequence 3, Application US/09210084

Patent No. 6197511

GENERAL INFORMATION:

APPLICANT: Hillman, Jennifer L.

APPLICANT: Lal, Preeti

TITLE OF INVENTION: NOVEL KALLIKREIN

NUMBER OF SEQUENCES: 5

CORRESPONDENCE ADDRESS:

ADDRESSEE: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Drive

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/210,084

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/824,874

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.

REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PF-0252 US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-855-0555

TELEFAX: 415-845-4166

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 253 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

LIBRARY: GenBank

CLONE: 532504

US-09-210-084-3

Query Match 100.0%; Score 50; DB 3; Length 253;

Best Local Similarity 100.0%; Pred. No. 0.16; Indels 0; Gaps 0;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY

1 GPLVCRGTL 9

|||||

207 GPLVCRGTL 215

DB

## RESULT 14

US-09-764-762-3

Sequence 3, Application US/09764762

Patent No. 6472195

GENERAL INFORMATION:

APPLICANT: Hillman, Jennifer L.

APPLICANT: Lal, Preeti

TITLE OF INVENTION: NOVEL KALLIKREIN

NUMBER OF SEQUENCES: 5

CORRESPONDENCE ADDRESS:

ADDRESSEE: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Drive

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/764,762

FILING DATE: 16-Jan-2001

CLASSIFICATION: &lt;Unknown&gt;

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/210,084

FILING DATE: &lt;Unknown&gt;

ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.

REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PF-0252 US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-855-0555



Db 207 GPLVCRGTL 215  
Search completed: March 1, 2004, 17:38:25  
Job time : 11.8889 secs

TELEFAX: 415-945-4166  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 253 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 532504  
SEQUENCE DESCRIPTION: SEQ ID NO: 3:  
3-09-764-762-3

Query Match 100.0%; Score 50; DB 4; Length 253;  
Best Local Similarity 100.0%; Pred. No. 0.16;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

y 1 GPLVCRGTL 9  
|||||  
b 207 GPLVCRGTL 215

RESULT 15  
CT-US96-04294-2  
Sequence 2, Application PC/TUS9604294  
GENERAL INFORMATION:  
APPLICANT: Dixon, Eric P.  
APPLICANT: Johnstone, Edward M.  
APPLICANT: Little, Sheila P.  
TITLE OF INVENTION: AMYLOID PRECURSOR PROTEIN PROTEASE AND  
RELATED NUCLEIC ACIDS  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESS: Eli Lilly and Company  
STREET: Lilly Corporate Center  
CITY: Indianapolis  
STATE: Indiana  
COUNTRY: United States of America  
ZIP: 46285  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/04294  
FILING DATE:

CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/416,257  
FILING DATE: 04-APR-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Blalock, Donna K.  
REGISTRATION NUMBER: 38,082  
REFERENCE/DOCKET NUMBER: X9239  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 317-277-1090  
TELEFAX: 317-276-3861  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 253 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
CT-US96-04294-2

Query Match 100.0%; Score 50; DB 5; Length 253;  
Best Local Similarity 100.0%; Pred. No. 0.16;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

y 1 GPLVCRGTL 9  
|||||

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

M protein - protein search, using sw model

un on: March 1, 2004, 17:16:55 ; Search time 45.5556 Seconds  
(without alignments)  
55,820 Million cell updates/sec

title: US-09-905-083-80

effect score: 50

sequence: 1 GPLVCRGTL 9

coring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

searched: 1586107 seqs, 282547505 residues

total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04.\*

- 1: Geneseq1980s.\*
- 2: Geneseq1990s.\*
- 3: Geneseq2000s.\*
- 4: Geneseq2001s.\*
- 5: Geneseq2002s.\*
- 6: Geneseq2003as.\*
- 7: Geneseq2003bs.\*
- 8: Geneseq2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	50	100.0	9	AAE08285	Aae08285 Human str
2	50	100.0	181	ADA05738	Ada05738 Human NOV
3	50	100.0	198	ADA05736	Ada05736 Human NOV
4	50	100.0	224	ADA05744	Ada05744 Human NOV
5	50	100.0	225	AAB98502	Aab98502 Human Str
6	50	100.0	247	ADA05742	Ada05742 Human NOV
7	50	100.0	250	ADA05732	Ada05732 Human NOV
8	50	100.0	252	ADA05734	Ada05734 Human NOV
9	50	100.0	253	AAR67888	Aar67888 Human str
10	50	100.0	253	AAMC5383	Aamc5383 Human amy
11	50	100.0	253	ABB84421	Abb84421 Human SCC
12	50	100.0	253	ABB84406	Abb84406 Human SCC
13	50	100.0	253	AAB82740	Aab82740 Amino aci
14	50	100.0	253	ABU07440	Abu07440 Protein d
15	50	100.0	253	ABU07471	Abu07471 Protein d
16	50	100.0	253	ARK58471	Ark58471 Human str
17	50	100.0	253	ABR80484	ABr80484 Ovarian c
18	50	100.0	257	AAAB21326	AAab21326 Human HSC
19	47	94.0	249	5 ABB84420	5 ABB84420 Porcine S
20	45	90.0	247	5 AAB86677	5 AAB86677 Novel hum
21	45	90.0	247	4 AAU23217	4 AAu23217 Novel hum
22	45	90.0	247	4 AAU23752	4 AAu23752 Novel hum
23	45	90.0	247	4 AAU17043	4 AAu17043 Human nov
24	45	90.0	247	7 ADB60011	7 ADB60011 Connectiv
25	45	90.0	250	3 AAB21298	3 AAB21298 Human KLK

26 45 90.0 250 5 ABP64969 Human pro  
27 45 90.0 250 6 ABR55400 Amino aci  
28 45 90.0 251 4 AAU16971 Human nov  
29 45 90.0 296 3 AAB21297 Human KLK  
30 44 88.0 237 3 AAB21305 Human KLK  
31 44 88.0 243 5 ABB84419 Bovine SC  
32 44 88.0 251 3 AAB21306 Human KLK  
33 44 88.0 267 6 ABJ26666 Human pro  
34 42 84.0 115 4 ABG19361 Novel hum  
35 41 82.0 23 5 ABB84409 Human SCC  
36 41 82.0 276 3 AAB11713 Mouse ser  
37 41 82.0 621 5 ABP66315 Bifidobac  
38 41 82.0 3201 4 AAB82214 Polyketid  
39 41 82.0 3546 4 AAB82212 Polyketid  
40 40 80.0 9 2 AAY48056 Immunogen  
41 40 80.0 15 2 AAW58064 Human pro  
42 40 80.0 15 2 AAW58063 Human pro  
43 40 80.0 20 2 AAW11019 Antigen p  
44 40 80.0 125 4 AAG75172 Human col  
45 40 80.0 194 6 ADA50560 Prostate

## ALIGNMENTS

## RESULT 1

AAE08285

ID AAE08285 standard; peptide; 9 AA.

XX AAE08285;

XX 01-NOV-2001 (first entry)

XX Human stratum corneum chymotrypsin enzyme peptide #50 (residues 207-215).

DE Stratum corneum chymotrypsin enzyme; SCCE; cytostatic; vaccine; tumour;  
KW cancer; ovarian; breast; lung; colon; prostate; carcinoma; human;  
KW antiseptic therapy; malignant hyperplasia.

XX Homo sapiens.

XX WC200159158-A1.

XX 16-AUG-2001.

XX 07-FEB-2001; 2001WO-US003977.

XX 11-FEB-2000; 2000US-00502600.

XX (UYAR-) UNIV ARKANSAS.

XX O'brien TJ;

XX WPI; 2001-514676/56.

XX Diagnosing cancer comprises detecting stratum corneum chymotrypsin enzyme.

XX Claim 25; Page 113; 127pp; English.

The invention relates to diagnosing cancer especially ovarian cancer, by screening for stratum corneum chymotrypsin enzyme (SCCE). Proteases are considered to be an integral part of tumour growth and metastasis, and therefore, markers indicative of their presence or absence are useful for the diagnosis of cancer. The method is useful for diagnosing cancer, carcinoma and malignant hyperplasia. The SCCE oligonucleotide may be used to treat a cancer selected from ovarian, breast, lung, colon, prostate and other cancers in which SCCE is overexpressed. The present sequence is human SCCE peptide

SQ Sequence 9 AA;

Query Match 100.0%; Score 50; DB 4; Length 9;

[illegible]

5 Homo sapiens.  
 X WO2003029424-A2.  
 N 10-APR-2003.  
 D  
 K  
 K  
 F 02-OCT-2002; 2002WO-US031373.  
 X  
 R 02-OCT-2001; 2001US-0326483P.  
 R 05-OCT-2001; 2001US-0327435P.  
 R 05-OCT-2001; 2001US-0327449P.  
 R 09-OCT-2001; 2001US-0327917P.  
 R 09-OCT-2001; 2001US-0328029P.  
 R 09-OCT-2001; 2001US-0328044P.  
 R 09-OCT-2001; 2001US-0328056P.  
 R 12-OCT-2001; 2001US-0328849P.  
 R 12-OCT-2001; 2001US-0329414P.  
 R 15-OCT-2001; 2001US-0330142P.  
 R 17-OCT-2001; 2001US-0330309P.  
 R 18-OCT-2001; 2001US-0330309P.  
 R 22-OCT-2001; 2001US-0341058P.  
 R 24-OCT-2001; 2001US-0339266P.  
 R 24-OCT-2001; 2001US-0343629P.  
 R 29-OCT-2001; 2001US-0349575P.  
 R 01-NOV-2001; 2001US-0346357P.  
 R 17-APR-2002; 2002US-0373260P.  
 R 19-APR-2002; 2002US-0373815P.  
 R 19-APR-2002; 2002US-0373817P.  
 R 19-APR-2002; 2002US-0373826P.  
 R 19-APR-2002; 2002US-0373884P.  
 R 22-APR-2002; 2002US-0374977P.  
 R 16-MAY-2002; 2002US-0381037P.  
 R 16-MAY-2002; 2002US-0381038P.  
 R 16-MAY-2002; 2002US-0381042P.  
 R 17-MAY-2002; 2002US-0381642P.  
 R 28-MAY-2002; 2002US-0383656P.  
 R 29-MAY-2002; 2002US-0383831P.  
 R 25-JUN-2002; 2002US-0391335P.  
 R 01-OCT-2002; 2002US-00262511.  
 X (CURA-) CURAGEN CORP.  
 X  
 X Smithson G, Millet I, Peyman JA, Kekuda R, Ju J, Li L, Guo X;  
 PI Patcurajan M, Sytek KA, Edinger SK, Ellerman K, Malyankar UM;  
 PI Ort T, Gorman L, Zernusen BD, Anderson DW, Zhong M, Catterton E;  
 PI Ji W, Miller CE, Rastelli L, Stone DJ, Pena CE, Shenoy SG;  
 PI Shinkens RA, Rottenberg MB, Leach MD, Agee ML, Berghs C, Dipippo VA;  
 PI Eisen AJ, Gangolli EA, Rieger DK, Spaderna SK;  
 X WPI: 2003-381626/36.  
 DR N-PSDB; ADA05735.  
 X  
 X New NOVX polypeptides and nucleic acids, useful for diagnosing,  
 PT preventing or treating NOVX-associated disorders, e.g. diabetes, obesity,  
 PT cancer or dyslipidemia, and in chromosome mapping, tissue typing or  
 PT pharmacogenomics.  
 X  
 X Claim 1; Page 170; 586pp; English.  
 PS  
 X The present invention describes NOVX proteins, where X can be 1 to 55  
 CC (e.g. NOVX). Also described: (1) a composition comprising a polypeptide  
 CC described above and a carrier; (2) a kit comprising, in one or more  
 CC containers, the composition described above; (3) an isolated nucleic acid  
 CC molecule which encodes a NOVX protein of the invention; (4) a vector  
 CC comprising the nucleic acid molecule described above; (5) a cell  
 CC comprising the above vector; (6) an antibody that immunospecifically  
 CC binds to the polypeptide described above; (7) methods for determining the  
 CC presence or amount of the above polypeptide or nucleic acid molecule in a  
 CC sample; (8) methods for determining the presence of or predisposition to  
 CC a disease associated with altered levels of expression of the above  
 CC polypeptide or nucleic acid molecule in a first mammalian subject; (9) a  
 CC method of identifying an agent that binds to the polypeptide described  
 CC above; (10) a method for identifying a potential therapeutic agent for  
 CC use in treating a pathology that is related to an aberrant expression or

CC aberrant physiological interactions of the polypeptide; (11) a method of  
 CC screening for a modulator of activity or of latency or predisposition to  
 CC a pathology associated with the polypeptide; (12) a method for modulating  
 CC the activity of the polypeptide described above; (13) methods of treating  
 CC or preventing a pathology associated with the above polypeptide in a  
 CC mammal; and (14) a method for producing the above polypeptide. NOVX  
 CC sequences have antidiabetic, anorectic, antibacterial, virucide,  
 CC immunomodulator, cyrostatic, nootropic, neuroprotective, antiparkinsonian  
 CC and antilipidemic activities, and can be used in gene therapy. The  
 CC polypeptide is useful in manufacturing a medicament for treating a  
 CC syndrome associated with a human disease. The polypeptide or the nucleic  
 CC acid molecule may be used to diagnose, treat or prevent metabolic  
 CC disorders such as diabetes or obesity, infections, cachexia, cancer,  
 CC neurodegenerative disorders such as Alzheimer's disease or Parkinson's  
 CC disease, immune disorders, haematopoietic disorders and various  
 CC dyslipidaemias. The nucleic acids can also be used as hybridisation  
 CC probes, in chromosome mapping, tissue typing, preventive medicine and  
 CC pharmacogenomics. The present sequence represents a human NOVX from the  
 CC present invention.  
 XX  
 SQ Sequence 198 AA;  
 Query Match 100.0%; Score 50; DB 6; Length 198;  
 Best Local Similarity 100.0%; Pred. No. 1.4;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GPLVCRGTL 9  
 Db 152 GPLVCRGTL 160  
 |||||  
 RESULT 4  
 ADA05744  
 ID ADA05744 standard; protein; 224 AA.  
 XX  
 AC ADA05744;  
 XX  
 DT 06-NOV-2003 (first entry)  
 XX  
 DE Human NOV18g protein SEQ ID NO:104.  
 XX  
 KW human; NOVX; antidiabetic; anorectic; antibacterial; virucide;  
 KW immunomodulator; cyrostatic; nootropic; neuroprotective;  
 KW antiparkinsonian; antilipidemic; gene therapy; human disease;  
 KW metabolic disorder; diabetes; obesity; infection; cachexia; cancer;  
 KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;  
 KW immune disorder; haematopoietic disorder; dyslipidaemia.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO2003029424-A2.  
 XX  
 PD 10-APR-2003.  
 XX  
 PF 02-OCT-2002; 2002WO-US031373.  
 XX  
 PR 02-OCT-2001; 2001US-0326483P.  
 PR 05-OCT-2001; 2001US-0327435P.  
 PR 05-OCT-2001; 2001US-0327449P.  
 PR 09-OCT-2001; 2001US-0327917P.  
 PR 09-OCT-2001; 2001US-0328029P.  
 PR 09-OCT-2001; 2001US-0328044P.  
 PR 09-OCT-2001; 2001US-0328056P.  
 PR 12-OCT-2001; 2001US-0328849P.  
 PR 15-OCT-2001; 2001US-0329414P.  
 PR 17-OCT-2001; 2001US-0330142P.  
 PR 18-OCT-2001; 2001US-0330309P.  
 PR 22-OCT-2001; 2001US-0341058P.  
 PR 24-OCT-2001; 2001US-0339266P.  
 PR 24-OCT-2001; 2001US-0343629P.  
 PR 29-OCT-2001; 2001US-0349575P.  
 PR 01-NOV-2001; 2001US-0346357P.  
 PR 17-APR-2002; 2002US-0373260P.

1 19-APR-2002; 2002US-0373815P.  
 2 19-APR-2002; 2002US-0373817P.  
 3 19-APR-2002; 2002US-0373826P.  
 4 19-APR-2002; 2002US-0373884P.  
 5 22-APR-2002; 2002US-0374977P.  
 6 16-MAY-2002; 2002US-0381037P.  
 7 16-MAY-2002; 2002US-0381038P.  
 8 16-MAY-2002; 2002US-0381042P.  
 9 17-MAY-2002; 2002US-0381642P.  
 10 28-MAY-2002; 2002US-0383656P.  
 11 29-MAY-2002; 2002US-0383831P.  
 12 25-JUN-2002; 2002US-0391335P.  
 13 01-OCT-2002; 2002US-00262511.

(CURA-) CURAGEN CORP.

Smithson G, Millet I, Peyman JA, Kekuda R, Ju J, Li L, Guo X;  
 Patturajan M, Spyrek KA, Edinger SR, Ellerman K, Malyankar UM;  
 Ort T, Gorman L, Zerkush BD, Anderson DW, Zhong M, Catterton E;  
 Ji W, Miller CE, Rastelli L, Stone DJ, Pena CE, Shenoy SG;  
 Shimkets RA, Rothenberg ME, Leach MD, Agee ML, Berghs C, Dipippo VA;  
 Eisen AJ, Gangolli EA, Rieger DK, Spaderna SK;

WPI; 2003-381626/36.

N-PSDB; ADA05743.

New NOVX polypeptides and nucleic acids, useful for diagnosing,  
 preventing or treating NOVX-associated disorders, e.g. diabetes, obesity,  
 cancer or dyslipidemia, and in chromosome mapping, tissue typing or  
 pharmacogenomics.

Claim 1; Page 172; 586pp; English.

The present invention describes NOVX proteins, where X can be 1 to 55  
 (e.g. NOV1). Also described: (1) a composition comprising a polypeptide  
 described above and a carrier; (2) a kit comprising, in one or more  
 containers, the composition described above; (3) an isolated nucleic acid  
 molecule which encodes a NOVX protein of the invention; (4) a vector  
 comprising the nucleic acid molecule described above; (5) a cell  
 comprising the above vector; (6) an antibody that immunospecifically  
 binds to the polypeptide described above; (7) methods for determining the  
 presence or amount of the above polypeptide or nucleic acid molecule in a  
 sample; (8) methods for determining the presence of or predisposition to  
 a disease associated with altered levels of expression of the above  
 polypeptide or nucleic acid molecule in a first mammalian subject; (9) a  
 method of identifying an agent that binds to the polypeptide described  
 above; (10) a method for identifying a potential therapeutic agent for  
 use in treating a pathology that is related to an aberrant expression or  
 aberrant physiological interactions of the polypeptide; (11) a method of  
 screening for a modulator of activity or of latency or predisposition to  
 a pathology associated with the polypeptide; (12) a method for modulating  
 the activity of the polypeptide described above; (13) methods of treating  
 or preventing a pathology associated with the above polypeptide in a  
 mammal; and (14) a method for producing the above polypeptide. NOVX  
 sequences have antidiabetic, anorectic, neuroprotective, antiparkinsonian  
 immunomodulator, cytostatic, nontropic, neuroprotective, antiviral,  
 and antilipemic activities, and can be used in gene therapy. The  
 polypeptide is useful in manufacturing a medicament for treating a  
 syndrome associated with a human disease. The polypeptide or the nucleic  
 acid molecule may be used to diagnose, treat or prevent metabolic  
 disorders such as diabetes or obesity, infections, cachexia, cancer,  
 neurodegenerative disorders such as Alzheimer's disease or Parkinson's  
 disease, immune disorders, haematopoietic disorders and various  
 dyslipidemias. The nucleic acids can also be used as hybridisation  
 probes, in chromosome mapping, tissue typing, preventive medicine and  
 pharmacogenomics. The present sequence represents a human NOVX from the  
 present invention.

Sequence 224 AA;

Query Match 100.0%; Score 50; DB 6; Length 224;  
 Best Local Similarity 100.0%; Pred. No. 1.6;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GPLVCRGTL 9  
 |||||  
 Db 193 GPLVCRGTL 201

# RESULT 5

AAB98502  
 ID AAB98502 standard; protein; 225 AA.

XX AC AAB98502;

XX XX 03-AUG-2001 (first entry)

XX XX Human Stratum Corneum Chymotryptic Enzyme, SCCE, catalytic domain.

XX XX Human; TADG-15; cytostatic; vaccine; ovarian tumour; cancer;

XX XX tumour antigen-derived gene 15; serine protease;

XX XX Stratum Corneum Chymotryptic Enzyme; SCCE.

XX OS Homo sapiens.

XX XX WO200129056-A1.

XX XX 26-APR-2001.

XX XX 20-OCT-2000; 2000WO-US029095.

XX XX 20-OCT-1999; 99US-00421213.

XX XX (UYAR-) UNIV ARKANSAS.

XX XX O'Brien TJ, Tanimoto H;

XX XX WPI; 2001-381031/40.

XX XX Novel extracellular serine protease, termed tumor antigen-derived gene 15

XX XX protein overexpressed in carcinomas and DNA encoding it, for diagnosis,

XX XX treatment, prevention of cancer, particularly breast, ovarian cancer.

XX XX Example 10; Fig 1; 130pp; English.

XX XX The present invention relates to human tumour antigen-derived gene 15

XX XX (TADG-15) protein and coding sequence (see AAB98501 and AAB98500). TADG-

XX XX 15 is an extracellular serine protease. It was found that TADG-15 is over

XX XX -expressed in ovarian tumours. TADG-15 protein or its fragments of 9-20

XX XX residues that lack TADG-15 protease activity are useful for vaccinating

XX XX an individual against TADG-15, having, suspected of having or at risk of

XX XX getting cancer. Furthermore, the TADG-15 gene can be used as a diagnostic

XX XX or therapeutic target in cancer. The present sequence was used in a

XX XX sequence homology alignment with the catalytic domain of TADG-15

XX XX SQ Sequence 225 AA;

Query Match 100.0%; Score 50; DB 4; Length 225;

Best Local Similarity 100.0%; Pred. No. 1.6;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GPLVCRGTL 9

|||||

Db 179 GPLVCRGTL 187

# RESULT 6

ADA05742  
 ID ADA05742 standard; protein; 247 AA.

XX AC ADA05742;

XX XX 06-NOV-2003 (first entry)

XX XX Human NOV18f protein SEQ ID NO:102.

human; NOVX; antidiabetic; anorectic; antibacterial; virucide;  
immunomodulator; cytosstatic; nootropic; neuroprotective;  
antiparkinsonian; antilipemic; gene therapy; human disease;  
metabolic disorder; diabetes; obesity; infection; cachexia; cancer;  
neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;  
immune disorder; haematopoietic disorder; dyslipidaemia.  
Homo sapiens.  
WO2003029424-A2.  
10-APR-2003.  
02-OCT-2002; 2002WO-US031373.  
02-OCT-2001; 2001US-0336483P.  
05-OCT-2001; 2001US-0327435P.  
05-OCT-2001; 2001US-0327449P.  
09-OCT-2001; 2001US-0327917P.  
09-OCT-2001; 2001US-0328029P.  
09-OCT-2001; 2001US-0328044P.  
09-OCT-2001; 2001US-0328056P.  
12-OCT-2001; 2001US-0328849P.  
15-OCT-2001; 2001US-0329414P.  
17-OCT-2001; 2001US-0330142P.  
18-OCT-2001; 2001US-0330309P.  
22-OCT-2001; 2001US-0341058P.  
24-OCT-2001; 2001US-0339266P.  
24-OCT-2001; 2001US-0343629P.  
29-OCT-2001; 2001US-0349575P.  
01-NOV-2001; 2001US-0346357P.  
17-APR-2002; 2002US-0373260P.  
19-APR-2002; 2002US-0373815P.  
19-APR-2002; 2002US-0373817P.  
19-APR-2002; 2002US-0373826P.  
22-APR-2002; 2002US-0373884P.  
22-APR-2002; 2002US-0374977P.  
26-MAY-2002; 2002US-0381037P.  
16-MAY-2002; 2002US-0381038P.  
17-MAY-2002; 2002US-0381042P.  
28-MAY-2002; 2002US-0381642P.  
29-MAY-2002; 2002US-0383656P.  
25-JUN-2002; 2002US-0383831P.  
01-OCT-2002; 2002US-0391335P.  
01-OCT-2002; 2002US-00262511.  
(CURA-) CUPAGEN CORP.  
Smithson G, Millet I, Peyman JA, Kekuda R, Ju J, Li L, Guo X;  
Patturajan M, Spytek KA, Edinger SR, Ellerman K, Malyankar UM;  
Ott T, Gorman L, Zernusen BD, Anderson DW, Zhong M, Catterton E;  
Ji W, Miller CE, Rastelli L, Stone DJ, Pena CEA, Shency SG;  
Shinkets RA, Rothenberg ME, Leach MD, Agee ML, Berghs C, Dipippo VA;  
Eisen AJ, Gangolli EA, Rieger DK, Spaderna SK;  
WPI; 2003-381526/36.  
N-PSDB; ADA05741.  
New NOVX polypeptides and nucleic acids, useful for diagnosing,  
preventing or treating NOVX-associated disorders, e.g. diabetes, obesity,  
cancer or dyslipidemia, and in chromosome mapping, tissue typing or  
pharmacogenomics.  
Claim 1; Page 172; 586pp; English.  
The present invention describes NOVX proteins, where X can be 1 to 55  
(e.g. NOV1). Also described: (1) a composition comprising a polypeptide  
described above and a carrier; (2) a kit comprising, in one or more  
containers, the composition described above; (3) an isolated nucleic acid  
molecule which encodes a NOVX protein of the invention; (4) a vector  
comprising the nucleic acid molecule described above; (5) a cell  
comprising the above vector; (6) an antibody that immunospecifically  
binds to the polypeptide described above; (7) methods for determining the

presence or amount of the above polypeptide or nucleic acid molecule in a  
sample; (8) methods for determining the presence of or predisposition to  
a disease associated with altered levels of expression of the above  
polypeptide or nucleic acid molecule in a first mammalian subject; (9) a  
method of identifying an agent that binds to the polypeptide described  
above; (10) a method for identifying a potential therapeutic agent for  
use in treating a pathology that is related to an aberrant expression or  
aberrant physiological interactions of the polypeptide; (11) a method of  
screening for a modulator of activity or of latency or predisposition to  
a pathology associated with the polypeptide; (12) a method for modulating  
the activity of the polypeptide described above; (13) methods of treating  
or preventing a pathology associated with the above polypeptide in a  
mammal; and (14) a method for producing the above polypeptide. NOVX  
sequences have antidiabetic, anorectic, antibacterial, virucide,  
immunomodulator, cytosstatic, nootropic, neuroprotective, antiparkinsonian  
and antilipemic activities, and can be used in gene therapy. The  
polypeptide is useful in manufacturing a medicament for treating a  
syndrome associated with a human disease. The polypeptide or the nucleic  
acid molecule may be used to diagnose, treat or prevent metabolic  
disorders such as diabetes or obesity, infections, cachexia, cancer,  
neurodegenerative disorders such as Alzheimer's disease or Parkinson's  
disease, immune disorders, haematopoietic disorders and various  
dyslipidaemias. The nucleic acids can also be used as hybridisation  
probes, in chromosome mapping, tissue typing, preventive medicine and  
pharmacogenomics. The present sequence represents a human NOVX from the  
present invention.  
SQ Sequence 247 AA;  
Query Match 100.0%; Score 50; DB 6; Length 247;  
Best Local Similarity 100.0%; Pred. No. 1.8; Mismatches 0; Indels 0; Gaps 0;  
Matches 9; Conservative 0;  
QY 1 GPLVCRGTL 9  
Db 193 GPLVCRGTL 201  
|||||  
RESULT 7  
ADA05732  
ID ADA05732 standard; protein; 250 AA.  
AC ADA05732;  
XX  
DT 06-NOV-2003 (first entry)  
XX  
DE Human NOV18a protein SEQ ID NO:92.  
XX  
KW human; NOVX; antidiabetic; anorectic; antibacterial; virucide;  
immunomodulator; cytosstatic; nootropic; neuroprotective;  
antiparkinsonian; antilipemic; gene therapy; human disease;  
metabolic disorder; diabetes; obesity; infection; cachexia; cancer;  
neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;  
immune disorder; haematopoietic disorder; dyslipidaemia.  
XX  
OS Homo sapiens.  
XX  
PN WO2003029424-A2.  
XX  
PD 10-APR-2003.  
XX  
PF 02-OCT-2002; 2002WO-US031373.  
XX  
PR 02-OCT-2001; 2001US-0326483P.  
PR 05-OCT-2001; 2001US-0327435P.  
PR 05-OCT-2001; 2001US-0327449P.  
PR 09-OCT-2001; 2001US-0327917P.  
PR 09-OCT-2001; 2001US-0328029P.  
PR 09-OCT-2001; 2001US-0328044P.  
PR 09-OCT-2001; 2001US-0328056P.  
PR 12-OCT-2001; 2001US-0328849P.  
PR 15-OCT-2001; 2001US-0329414P.  
PR 17-OCT-2001; 2001US-0330142P.

18-OCT-2001; 2001US-0330309P.  
22-OCT-2001; 2001US-0341058P.  
24-OCT-2001; 2001US-0339266P.  
24-OCT-2001; 2001US-0343629P.  
29-OCT-2001; 2001US-0349575P.  
01-NOV-2001; 2001US-0346357P.  
17-APR-2002; 2002US-0373260P.  
19-APR-2002; 2002US-0373815P.  
19-APR-2002; 2002US-0373817P.  
19-APR-2002; 2002US-0373826P.  
19-APR-2002; 2002US-0373884P.  
22-APR-2002; 2002US-0374977P.  
16-MAY-2002; 2002US-0381037P.  
16-MAY-2002; 2002US-0381038P.  
16-MAY-2002; 2002US-0381042P.  
28-MAY-2002; 2002US-0381642P.  
28-MAY-2002; 2002US-0383656P.  
29-MAY-2002; 2002US-0383811P.  
25-JUN-2002; 2002US-0391335P.  
01-OCT-2002; 2002US-00262511.

(CURA-) CURAGEN CORP.

Smithson G, Millet I, Peyman JA, Kekuda R, Ju J, Li L, Guo X;  
Patturajan M, Spytek KA, Edinger SR, Ellerman K, Malyankar UM;  
Ort I, Gorman L, Zerkusen BD, Anderson DW, Zhong M, Catterton E;  
Ji W, Miller CE, Raschelli L, Stone DU, Pena CEA, Shenoy SG;  
Shankets RA, Rothenberg ME, Leach MD, Agee ML, Berghs C, Dipippo VA;  
Eisen AJ, Gangolli EA, Rieger DK, Spaderna SK;

WPI; 2003-381626/36.

N-PSDB; ADA05731.

New NOVX polypeptides and nucleic acids, useful for diagnosing,  
preventing or treating NOVX-associated disorders, e.g. diabetes, obesity,  
cancer or dyslipidemia, and in chromosome mapping, tissue typing or  
pharmacogenomics.

Claim 1; Page 169-170; 586pp; English.

The present invention describes NOVX proteins, where X can be 1 to 55  
(e.g. NOVX). Also described: (1) a composition comprising a polypeptide  
described above and a carrier; (2) a kit comprising, in one or more  
containers, the composition described above; (3) an isolated nucleic acid  
molecule which encodes a NOVX protein of the invention; (4) a vector  
comprising the nucleic acid molecule described above; (5) a cell  
comprising the above vector; (6) an antibody that immunospecifically  
binds to the polypeptide described above; (7) methods for determining the  
presence or amount of the above polypeptide or nucleic acid molecule in a  
sample; (8) methods for determining the presence of or predisposition to  
a disease associated with altered levels of expression of the above  
polypeptide or nucleic acid molecule in a first mammalian subject; (9) a  
method of identifying an agent that binds to the polypeptide described  
above; (10) a method for identifying a potential therapeutic agent for  
use in treating a pathology that is related to an aberrant expression or  
aberrant physiological interactions of the polypeptide; (11) a method of  
screening for a modulator of activity or of latency or predisposition to  
a pathology associated with the polypeptide; (12) a method for modulating  
the activity of the polypeptide described above; (13) methods of treating  
or preventing a pathology associated with the above polypeptide in a  
mammal; and (14) a method for producing the above polypeptide. NOVX  
sequences have antidiabetic, anorectic, antibacterial, virucide,  
immunomodulator, cytostatic, neurotropic, neuroprotective, antiparkinsonian  
and antileptin activities, and can be used in gene therapy. The  
polypeptide is useful in manufacturing a medicament for treating a  
syndrome associated with a human disease. The polypeptide or the nucleic  
acid molecule may be used to diagnose, treat or prevent metabolic  
disorders such as diabetes or obesity, infectious, cachexia, cancer,  
neurodegenerative disorders such as Alzheimer's disease or Parkinson's  
disease, immune disorders, haematopoietic disorders and various  
dyslipidaemias. The nucleic acids can also be used as hybridisation  
probes, in chromosome mapping, tissue typing, preventive medicine and  
pharmacogenomics. The present sequence represents a human NOVX from the

CC present invention.

XX Sequence 250 AA;

Query Match 100.0%; Score 50; DB 6; Length 250;  
Best Local Similarity 100.0%; Pred. No. 1.8; Mismatches 0; Gaps 0;  
Matches 9; Conservative 0;

Qy 1 GPLVCRGTL 9  
Db 204 GPLVCRGTL 212  
|||||

RESULT 8

ADA05734  
ID ADA05734 standard; protein; 252 AA.

XX ADA05734;

XX 06-NOV-2003 (first entry)

XX Human NOV18b protein SEQ ID NO:94.

XX human; NOVX; antidiabetic; anorectic; antibacterial; virucide;  
XX immunomodulator; cytostatic; neurotropic; neuroprotective;  
XX antiparkinsonian; antileptin; gene therapy; human disease;  
XX metabolic disorder; diabetes; obesity; infection; cachexia; cancer;  
XX neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;  
XX immune disorder; haematopoietic disorder; dyslipidaemia.

XX Homo sapiens.

XX W02003029424-A2.

XX 10-APR-2003.

XX 02-OCT-2002; 2002WO-US031373.

XX 02-OCT-2001; 2001US-0326483P.

XX 05-OCT-2001; 2001US-0327435P.

XX 09-OCT-2001; 2001US-0327449P.

XX 09-OCT-2001; 2001US-0327917P.

XX 09-OCT-2001; 2001US-0328029P.

XX 09-OCT-2001; 2001US-0328044P.

XX 12-OCT-2001; 2001US-0328056P.

XX 15-OCT-2001; 2001US-032849P.

XX 17-OCT-2001; 2001US-0329414P.

XX 18-OCT-2001; 2001US-0330142P.

XX 22-OCT-2001; 2001US-0330309P.

XX 24-OCT-2001; 2001US-0341058P.

XX 24-OCT-2001; 2001US-0339266P.

XX 29-OCT-2001; 2001US-0343629P.

XX 01-NOV-2001; 2001US-0349575P.

XX 17-APR-2002; 2002US-0346357P.

XX 19-APR-2002; 2002US-0373260P.

XX 19-APR-2002; 2002US-0373815P.

XX 19-APR-2002; 2002US-0373817P.

XX 19-APR-2002; 2002US-0373826P.

XX 19-APR-2002; 2002US-0373884P.

XX 22-APR-2002; 2002US-0374977P.

XX 16-MAY-2002; 2002US-0381037P.

XX 16-MAY-2002; 2002US-0381038P.

XX 17-MAY-2002; 2002US-0381042P.

XX 28-MAY-2002; 2002US-0383656P.

XX 29-MAY-2002; 2002US-0383811P.

XX 25-JUN-2002; 2002US-0391335P.

XX 01-OCT-2002; 2002US-00262511.

XX (CURA-) CURAGEN CORP.

XX Smithson G, Millet I, Peyman JA, Kekuda R, Ju J, Li L, Guo X;

XX Patturajan M, Spytek KA, Edinger SR, Ellerman K, Malyankar UM;

I Ort T, Gorman L, Zerhusen BD, Anderson DW, Zhong M, Catterton E;  
 J Ji W, Miller CE, Rastelli L, Stone DU, Pena CE, Sheny SG;  
 I Shimkets RA, Rothenberg ME, Leach MD, Agee ML, Berghs C, Dipippo VA;  
 I Eisen AJ, Gangolli EA, Rieger DK, Spaderna SK;  
 K WPI; 2003-381626/36.  
 R N-PSDB; ADA05733.  
 R  
 X New NOVX polypeptides and nucleic acids, useful for diagnosing,  
 T preventing or treating NOVX-associated disorders, e.g. diabetes, obesity,  
 T cancer or dyslipidemia, and in chromosome mapping, tissue typing or  
 T pharmacogenomics.  
 X  
 X Claim 1; Page 170; 586pp; English.  
 X  
 C The present invention describes NOVX proteins, where X can be 1 to 55  
 C (e.g. NOV1). Also described: (1) a composition comprising a polypeptide  
 C described above and a carrier; (2) a kit comprising, in one or more  
 C containers, the composition described above; (3) an isolated nucleic acid  
 C molecule which encodes a NOVX protein of the invention; (4) a vector  
 C comprising the nucleic acid molecule described above; (5) a cell  
 C comprising the above vector; (6) an antibody that immunospecifically  
 C binds to the polypeptide described above; (7) methods for determining the  
 C presence or amount of the above polypeptide or nucleic acid molecule in a  
 C sample; (8) methods for determining the presence of or predisposition to  
 C a disease associated with altered levels of expression of the above  
 C polypeptide or nucleic acid molecule in a first mammalian subject; (9) a  
 C method of identifying an agent that binds to the polypeptide described  
 C above; (10) a method for identifying a potential therapeutic agent for  
 C use in treating a pathology that is related to an aberrant expression or  
 C aberrant physiological interactions of the polypeptide; (11) a method of  
 C screening for a modulator of activity or of latency or predisposition to  
 C a pathology associated with the polypeptide; (12) a method for modulating  
 C the activity of the polypeptide described above; (13) methods of treating  
 C or preventing a pathology associated with the above polypeptide in a  
 C mammal; and (14) a method for producing the above polypeptide. NOVX  
 C sequences have antidiabetic, anorectic, antibacterial, virucide,  
 C immunomodulator, cytosstatic, neurotropic, neuroprotective, antiparkinsonian  
 C and antilipemic activities, and can be used in gene therapy. The  
 C polypeptide is useful in manufacturing a medicament for treating a  
 C syndrome associated with a human disease. The polypeptide or the nucleic  
 C acid molecule may be used to diagnose, treat or prevent metabolic  
 C disorders such as diabetes or obesity, infections, cachexia, cancer,  
 C neurodegenerative disorders such as Alzheimer's disease or Parkinson's  
 C disease, immune disorders, haematopoietic disorders and various  
 C dyslipidaemias. The nucleic acids can also be used as hybridisation  
 C probes, in chromosome mapping, tissue typing, preventive medicine and  
 C pharmacogenomics. The present sequence represents a human NOVX from the  
 C present invention.  
 X  
 X Sequence 252 AA;  
 SQ  
 Query Match 100.0%; Score 50; DB 6; Length 252;  
 Best Local Similarity 100.0%; Pred. No. 1.8;  
 Matches .9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GPLVCRGTL 9  
 Db 206 GPLVCRGTL 214  
 |||||  
 RESULT 9  
 AAR67888  
 ID AAR67888 standard; protein; 253 AA.  
 XX  
 AC AAR67888;  
 XX  
 XX 25-MAR-2003 (revised)  
 DT 09-AUG-1995 (first entry)  
 XX  
 DE Human stratum corneum chymotryptic recombinant enzyme (SCCE).  
 XX Stratum corneum chymotryptic enzyme; skin disorder; acne; psoriasis;  
 KW

KW callosities; keratosis pilaris; ichthyoses; eczema.  
 XX  
 OS Homo sapiens.  
 XX WO9500651-A1.  
 PN  
 XX OS-JAN-1995.  
 PD  
 XX 20-JUN-1994; 94WO-IB000166.  
 PF  
 XX 18-JUN-1993; 93DK-00000725.  
 PR  
 XX (SYMB-) SYMBICOM AB.  
 PA  
 XX Egelrud T, Hansson L;  
 PI  
 XX WPI; 1995-052086/07.  
 DR  
 XX N-PSDB; RAQ81203.  
 XX  
 PT Nucleotide sequences encoding stratum corneum chymotryptic enzyme - and  
 PT related vectors, transformed cells and polypeptides, useful for treating  
 PT skin disorders, e.g. acne or psoriasis, and for identification of  
 PT specific inhibitors.  
 XX  
 PS Disclosure; Page 97; 137pp; English.  
 XX  
 CC The enzyme encoded by this sequence is used in pharmaceutical, cosmetic  
 CC and skin care products, especially to treat and prevent acne, xeroderma,  
 CC or other hyperkeratotic conditions (e.g. callosities or keratosis  
 CC pilaris), ichthyoses, psoriasis, eczema, etc. It is produced  
 CC recombinantly following mammal, insect, plant, or microorganism  
 CC transformation with plasmid pS507. (Updated on 25-MAR-2003 to correct PN  
 CC field.)  
 XX  
 SQ Sequence 253 AA;  
 Query Match 100.0%; Score 50; DB 2; Length 253;  
 Best Local Similarity 100.0%; Pred. No. 1.8;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GPLVCRGTL 9  
 Db 207 GPLVCRGTL 215  
 |||||  
 RESULT 10  
 AAW05383  
 ID AAW05383 standard; protein; 253 AA.  
 XX  
 AC AAW05383;  
 XX  
 XX 31-DEC-1996 (first entry)  
 DT  
 XX Human amyloid precursor protein protease.  
 DE  
 XX Amyloid precursor protein protease; Alzheimer's disease; diagnosis;  
 KW therapy.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO9631122-A1.  
 PD  
 XX 10-OCT-1996.  
 PF  
 XX 02-APR-1996; 96WO-US004294.  
 PR  
 XX 04-APR-1995; 95US-00416257.  
 XX  
 PA (ELIL ) LILLY & CO ELI.  
 XX  
 XX Dixon EP, Johnstone EM, Little SP;  
 PI  
 XX WPI; 1996-464694/46.  
 DR



CC for the study of disease with the aim of improving treatment, to relieve  
CC or ameliorate a pathogenic condition, for development or testing of a  
CC cosmetic or a pharmaceutical formulation, and for the development of a  
CC diagnostic method. It can also be used as a model for a skin disease or  
CC skin cancer. The invention is also useful for screening or identifying a  
CC compound or composition effective for the prevention or treatment of an  
CC abnormal or unwanted phenotype, and for screening or identifying a  
CC compound or composition effective for the prevention or treatment of  
CC inflammatory skin diseases selected from diseases consisting of epidermal  
CC hyperkeratosis, acanthosis, epidermal inflammation, dermal inflammation,  
CC pruritus, atopic dermatitis, eczema, acne and inherited skin diseases  
CC with epidermal hyperkeratosis. The mammal of the invention is also useful  
CC as a model for further studies of itch mechanisms and the testing of  
CC potential compounds and compositions for relieve of various skin diseases  
CC where itch is a component. This sequence represents the N-terminal  
CC fragment of the human stratum corneum chymotryptic enzyme, SCCE  
CC synonymous with human kallikrein 7 (KLK7), used in the development of the  
CC transgenic mammals described in the invention  
CC  
CC  
XX Sequence 253 AA;  
SQ  
Query Match 100.0%; Score 50; DB 5; Length 253;  
Best Local Similarity 100.0%; Pred. No. 1.8;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Y 1 GPLVCRGTL 9  
b 207 GPLVCRGTL 215  
RESULT 12  
BB84421  
D ABB84421 standard; peptide; 253 AA.  
X C ABB84421;  
X T 08-NOV-2002. (first entry)  
X E Human SCCE protein N-terminal fragment SEQ ID 48.  
X W SCCE; human; stratum corneum chymotryptic enzyme; kallikrein 7;  
X W serine protease; transgenic mammal; skin; skin disease; skin cancer;  
X W hyperkeratosis; acanthosis; epidermal inflammation; dermal inflammation;  
X W pruritus; atopic dermatitis; eczema; acne; itch; KLK7; enzyme.  
X W Homo sapiens.  
X S  
X N WO200262135-A2.  
X D 15-AUG-2002.  
X F 08-FEB-2002; 2002WO-IB001300.  
X X 09-FEB-2001; 2001CA-02332655.  
X R 09-FEB-2001; 2001DK-00000218.  
X A (EGEL/) EGELRUD T.  
X A (HANS/) HANSSON L.  
X I Egelrud T, Hansson L;  
X R WPI; 2002-643380/69.  
X T Transgenic mammal or its embryo useful as model for human disease, has  
X T heterologous nucleotide sequence coding for stratum corneum chymotryptic  
X T enzyme operably linked to promoter that drives its expression in skin.  
X S Example 6; Page 37; 74pp; English.  
X C This invention describes a novel non-human transgenic mammal or mammalian  
X C embryo having integrated within its genome, a heterologous nucleotide  
X C sequence comprising at least a significant part of a nucleotide sequence  
X C coding for a stratum corneum chymotryptic enzyme (SCCE) or its variant,  
X C operably linked to a promoter that drives expression of heterologous scce  
X C or its variant in skin. The product of the invention is useful as a model

CC for the study of disease with the aim of improving treatment, to relieve  
CC or ameliorate a pathogenic condition, for development or testing of a  
CC cosmetic or a pharmaceutical formulation, and for the development of a  
CC diagnostic method. It can also be used as a model for a skin disease or  
CC skin cancer. The invention is also useful for screening or identifying a  
CC compound or composition effective for the prevention or treatment of an  
CC abnormal or unwanted phenotype, and for screening or identifying a  
CC compound or composition effective for the prevention or treatment of  
CC inflammatory skin diseases selected from diseases consisting of epidermal  
CC hyperkeratosis, acanthosis, epidermal inflammation, dermal inflammation,  
CC pruritus, atopic dermatitis, eczema, acne and inherited skin diseases  
CC with epidermal hyperkeratosis. The mammal of the invention is also useful  
CC as a model for further studies of itch mechanisms and the testing of  
CC potential compounds and compositions for relieve of various skin diseases  
CC where itch is a component. This sequence represents the N-terminal  
CC fragment of the human stratum corneum chymotryptic enzyme, SCCE  
CC synonymous with human kallikrein 7 (KLK7), used in the development of the  
CC transgenic mammals described in the invention  
CC  
CC  
XX Sequence 253 AA;  
SQ  
Query Match 100.0%; Score 50; DB 5; Length 253;  
Best Local Similarity 100.0%; Pred. No. 1.8;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Y 1 GPLVCRGTL 9  
b 207 GPLVCRGTL 215  
RESULT 12  
BB84406  
ID ABB84406 standard; protein; 253 AA.  
X C ABB84406;  
X X 08-NOV-2002 (first entry)  
X X Human SCCE protein.  
X X SCCE; human; stratum corneum chymotryptic enzyme; kallikrein 7;  
X X serine protease; transgenic mammal; skin; skin disease; skin cancer;  
X X hyperkeratosis; acanthosis; epidermal inflammation; dermal inflammation;  
X X pruritus; atopic dermatitis; eczema; acne; itch; KLK7; enzyme.  
X X Homo sapiens.  
X OS  
X X WO200262135-A2.  
X X 15-AUG-2002.  
X X 08-FEB-2002; 2002WO-IB001300.  
X X 09-FEB-2001; 2001CA-02332655.  
X X 09-FEB-2001; 2001DK-00000218.  
X X (EGEL/) EGELRUD T.  
X X (HANS/) HANSSON L.  
X X Egelrud T, Hansson L;  
X X WPI; 2002-643380/69.  
X X Transgenic mammal or its embryo useful as model for human disease, has  
X X heterologous nucleotide sequence coding for stratum corneum chymotryptic  
X X enzyme operably linked to promoter that drives its expression in skin.  
X X Claim 10; Page 58-59; 74pp; English.  
X X This invention describes a novel non-human transgenic mammal or mammalian  
X X embryo having integrated within its genome, a heterologous nucleotide  
X X sequence comprising at least a significant part of a nucleotide sequence  
X X coding for a stratum corneum chymotryptic enzyme (SCCE) or its variant,  
X X operably linked to a promoter that drives expression of heterologous scce  
X X or its variant in skin. The product of the invention is useful as a model

coding for a stratum corneum chymotryptic enzyme (SCCE) or its variant, operably linked to a promoter that drives expression of heterologous scce or its variant in skin. The product of the invention is useful as a model for the study of disease with the aim of improving treatment, to relieve or ameliorate a pathogenic condition, for development or testing of a cosmetic or a pharmaceutical formulation, and for the development of a diagnostic method. It can also be used as a model for a skin disease or skin cancer. The invention is also useful for screening or identifying a compound or composition effective for the prevention or treatment of an abnormal or unwanted phenotype, and for screening or identifying a compound or composition effective for the prevention or treatment of inflammatory skin diseases selected from diseases consisting of epidermal hyperkeratosis, acanthosis, epidermal inflammation, dermal inflammation, pruritus, atopic dermatitis, eczema, acne and inherited skin diseases with epidermal hyperkeratosis. The mammal of the invention is also useful as a model for further studies of itch mechanisms and the testing of potential compounds and compositions for relieve of various skin diseases where itch is a component. This sequence represents the human stratum corneum chymotryptic enzyme, SCCE which is a serine protease synonymous with human kallikrein 7 (KUK7) and is used in the development of the transgenic mammals described in the invention

Q Sequence 253 AA;  
Query Match 100.0%; Score 50; DB 5; Length 253;  
Best Local Similarity 100.0%; Pred. No. 1.8; 0; Gaps 0;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

NY 1 GPLVCRGTL 9  
|||||  
b 207 GPLVCRGTL 215

RESULT 13  
AAU82740  
D AAU82740 standard; protein; 253 AA.

AC AAU82740;

DT 23-APR-2002 (first entry)

DE Amino acid sequence of novel human protease #39.

Human; protease; cancer; immune-related disorder; cardiovascular disease; neuronal-associated disease; metabolic disorder; inflammatory disorder; nervous system disorder; sexual dysfunction; pain; mood disorder; hypertension; psychotic disorder; neurological disorder; dyskinesia; viral infection; human immunodeficiency virus; HIV; non-viral infection; ocular disease; cytostatic; enzyme.

OS Homo sapiens.

PN WO200200860-A2.

PD 03-JAN-2002.

PF 26-JUN-2001; 2001WO-US020171.

PR 26-JUN-2000; 2000US-0214047P.

PA (SUGE-) SUGEN INC.

PI Plowman G, Whyte D, Sudarsanam S, Manning G, Caenepeel S;  
PI Charydzak G;

DR WPI; 2002-139913/18.

DR N-PSDB; ABK31782.

XX Nucleic acids encoding novel human proteases, useful for useful for treating diseases and disorders such as cancers, immune-related diseases and disorders, cardiovascular disease (e.g. restenosis) and inflammatory disorders.

PS Claim 6; Fig 2N; 313pp; English.

XX The present invention relates to the isolation of novel human proteases, and the nucleic acids encoding them. The sequences of the invention are useful for treating diseases and disorders such as cancers (e.g. breast, colon, lung), immune-related diseases and disorders (e.g. inflammatory diseases and asthma), cardiovascular diseases (e.g. restenosis and coronary thrombosis), brain or neuronal-associated diseases, metabolic disorders (e.g. diabetes, obesity), inflammatory disorders (e.g. rheumatoid arthritis and psoriasis), central or peripheral nervous system diseases, migraines, pain, sexual dysfunction, mood disorders, attention disorders, cognitive disorders, hypertension, psychotonic disorders, neurological disorders (e.g. Alzheimer's disease, Parkinson's disease) and dyskinesias. The nucleic acids and polypeptides are also useful for treating viral infections caused by human immunodeficiency virus (HIV), and non-viral infections such as ocular disease (e.g. glaucoma) and macular degeneration. AAU82702-AAU82760 represent the novel human proteases of the invention

XX Sequence 253 AA;

Query Match 100.0%; Score 50; DB 5; Length 253;  
Best Local Similarity 100.0%; Pred. No. 1.8;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9  
|||||  
Db 207 GPLVCRGTL 215

RESULT 14  
ABU07440

ID ABU07440 standard; protein; 253 AA.

AC ABU07440;

DT 28-JAN-2003 (first entry)

DE Protein differentially regulated in prostate cancer #43:

Prostate cancer; gene expression; differential regulation; molecular marker; drug target; cancer detection; cancer diagnosis; cancer staging; cancer grading; cancer assessing; cancer monitoring.

OS Homo sapiens.

PN WO200281638-A2.

PD 17-OCT-2002.

PF 08-APR-2002; 2002WO-US010824.

PR 06-APR-2001; 2001US-0281731P.

PR 06-APR-2001; 2001US-0281732P.

PA (ORIG-) ORIGENE TECHNOLOGIES INC.

PI Sun Z, Jay G;

DR WPI; 2003-058520/05.

DR N-PSDB; ABX10343.

XX Novel genes which are differentially regulated in prostate cancer, useful for diagnosing prostate cancer in prostate tissue sample and assessing therapeutic or preventive intervention in prostate cancer patients.

PS Claim 1; Page 293-294; 416pp; English.

XX The invention describes genes (I) which are differentially regulated in prostate cancer. (I) is useful for diagnosing a prostate cancer in a sample comprising prostate tissue, which involves determining the number of target genes which are differentially-regulated in the sample, where the number is indicative of the probability that the sample comprises

prostate cancer. (I) Is useful for assessing a therapeutic or preventive intervention in a subject having a prostate cancer, which involves determining the expression levels in a sample comprising prostate tissue of target genes which are differentially-regulated in prostate cancer. Preferably, the expression levels of at least 10 genes are determined. (I) is also useful for identifying agents that modulate a biological activity of a polypeptide differentially-regulated in prostate cancer cells, which involves contacting a polypeptide differentially-regulated in prostate cancer cells with a test agent under conditions effective for the test agent to modulate a biological activity of the polypeptide, and determining whether the test agent modulates the biological activity. (I) is useful as molecular markers, as drug targets, and for detecting, diagnosing, staging, grading, assessing, monitoring, prognosticating, preventing or treating, determining predisposition to diseases and conditions especially relating to prostate cancer. (I) and its expression products are used in the diagnostic test to assay for presence of cancer e.g., in tissue sections, in biopsy sample, in total RNA, in lymph, in blood etc. (I) is useful for assessing cancer e.g., to determine the type of cancer, its stage of development, the nature of genetic defect, etc. The polypeptide encoded by (I) can be used as target for therapy or drug discovery. (I) can also be used for expressing the polypeptide and thus for searching specific binding partners of the polypeptide. (I) is useful in therapeutic applications to treat prostate cancer. The identification of specific genes, and groups of genes, expressed in pathways physiologically relevant to prostate cancer permits the definition of functional and disease pathways and the delineation of targets in these pathways which are useful in diagnostic, therapeutic, and clinical applications. This is the amino acid sequence of a protein differentially regulated in prostate cancer

Sequence 253 AA;

Query Match 100.0%; Score 50; DB 6; Length 253;

Best Local Similarity 100.0%; Pred. No. 1.8;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 GPLVCRGTL 9

b 207 GPLVCRGTL 215

RESULT 15

B007471

D ABU07471 standard; protein; 253 AA.

X C ABU07471;

X T 28-JAN-2003 (first entry)

X E Protein differentially regulated in prostate cancer #74.

X W Prostate cancer; gene expression; differential regulation;

X M molecular marker; drug target; cancer detection; cancer diagnosis;

X W cancer staging; cancer grading; cancer assessing; cancer monitoring.

X X Homo sapiens.

X US NC020281638-A2.

X N 17-OCT-2002.

X D 08-APR-2002; 2002WO-US010824.

X F 06-APR-2001; 2001US-0281731P.

X R 06-APR-2001; 2001US-0281732P.

X X (ORIG-) ORIGENE TECHNOLOGIES INC.

X A Sun Z, Jay G;

X I WPI; 2003-058520/05.

X R N-PSDB; ABX10375.

Novel genes which are differentially regulated in prostate cancer, useful for diagnosing prostate cancer in prostate tissue sample and assessing therapeutic or preventive intervention in prostate cancer patients.

Claim 1; Page 351; 416pp; English.

The invention describes genes (I) which are differentially regulated in prostate cancer. (I) is useful for diagnosing a prostate cancer in a sample comprising prostate tissue, which involves determining the number of target genes which are differentially-regulated in the sample, where the number is indicative of the probability that the sample comprises prostate cancer. (I) is useful for assessing a therapeutic or preventive intervention in a subject having a prostate cancer, which involves determining the expression levels in a sample comprising prostate tissue of target genes which are differentially-regulated in prostate cancer. Preferably, the expression levels of at least 10 genes are determined. (I) is also useful for identifying agents that modulate a biological activity of a polypeptide differentially-regulated in prostate cancer cells, which involves contacting a polypeptide differentially-regulated in prostate cancer cells with a test agent under conditions effective for the test agent to modulate a biological activity of the polypeptide, and determining whether the test agent modulates the biological activity. (I) is useful as molecular markers, as drug targets, and for detecting, diagnosing, staging, grading, assessing, monitoring, prognosticating, preventing or treating, determining predisposition to diseases and conditions especially relating to prostate cancer. (I) and its expression products are used in the diagnostic test to assay for presence of cancer e.g., in tissue sections, in biopsy sample, in total RNA, in lymph, in blood etc. (I) is useful for assessing cancer e.g., to determine the type of cancer, its stage of development, the nature of genetic defect, etc. The polypeptide encoded by (I) can be used as target for therapy or drug discovery. (I) can also be used for expressing the polypeptide and thus for searching specific binding partners of the polypeptide. (I) is useful in therapeutic applications to treat prostate cancer. The identification of specific genes, and groups of genes, expressed in pathways physiologically relevant to prostate cancer permits the definition of functional and disease pathways and the delineation of targets in these pathways which are useful in diagnostic, therapeutic, and clinical applications. This is the amino acid sequence of a protein differentially regulated in prostate cancer

Sequence 253 AA;

Query Match 100.0%; Score 50; DB 6; Length 253;

Best Local Similarity 100.0%; Pred. No. 1.8;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GPLVCRGTL 9

Db 207 GPLVCRGTL 215

Search completed: March 1, 2004, 17:28:55

Job time : 47.5556 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

M protein - protein search, using sw model

run on: March 1, 2004, 17:35:01 ; Search time 24.1111 Seconds  
(without alignments)  
78.818 Million cell updates/sec

Title: US-09-905-083-80

Effect score: 50

Sequence: 1 GPLVCRGTL 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 809742 seqs, 211153259 residues

Total number of hits satisfying chosen parameters: 809742

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

Published Applications\_AA:\*  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09D\_PUBCOMB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	50	100.0	9	9	US-09-918-243-80
2	50	100.0	9	9	US-09-905-083-80
3	50	100.0	253	9	US-09-888-615-98
4	50	100.0	253	14	US-09-764-762-3
5	50	100.0	253	15	US-10-264-283-90
6	50	100.0	253	15	US-10-295-027-498
7	50	100.0	253	15	US-10-173-999-48
8	45	90.0	247	9	US-09-764-898-284
9	45	90.0	247	9	US-09-764-847-742
10	45	90.0	247	14	US-10-052-154-742
11	45	90.0	250	16	US-10-055-569A-69
12	45	90.0	251	9	US-09-768-898-212
13	41	82.0	249	16	US-10-055-569A-68
14	41	82.0	276	16	US-10-055-569A-67
15	40	80.0	54	10	US-09-860-739-5

15 40 80.0 125 14 US-10-106-698-5946 Sequence 5946, App  
17 40 80.0 218 15 US-10-115-479-12 Sequence 12, Appl  
18 40 80.0 237 10 US-09-860-739-1 Sequence 1, Appl  
19 40 80.0 237 14 US-10-128-966-1 Sequence 1, Appl  
20 40 80.0 237 14 US-10-128-966-7 Sequence 7, Appl  
21 40 80.0 237 14 US-10-128-966-8 Sequence 8, Appl  
22 40 80.0 244 14 US-10-128-966-5 Sequence 5, Appl  
23 40 80.0 261 9 US-09-796-294-10 Sequence 10, Appl  
24 40 80.0 261 9 US-09-822-827-946 Sequence 946, App  
25 40 80.0 261 9 US-09-863-896-6 Sequence 6, Appl  
26 40 80.0 261 9 US-09-907-402-1 Sequence 1, Appl  
27 40 80.0 261 9 US-09-895-793-946 Sequence 946, App  
28 40 80.0 261 14 US-10-131-241-62 Sequence 62, Appl  
29 40 80.0 261 14 US-10-205-823-214 Sequence 214, App  
30 40 80.0 261 14 US-10-128-966-3 Sequence 3, Appl  
31 40 80.0 261 14 US-10-461-787-10 Sequence 10, Appl  
32 40 80.0 261 15 US-10-298-965-18 Sequence 18, Appl  
33 40 80.0 261 15 US-10-298-965-20 Sequence 20, Appl  
34 40 80.0 261 15 US-10-012-697-1558 Sequence 1558, App  
35 40 80.0 261 15 US-10-341-434-34 Sequence 34, Appl  
36 40 80.0 261 15 US-10-341-434-41 Sequence 41, Appl  
37 40 80.0 261 15 US-10-117-937-78 Sequence 78, Appl  
38 40 80.0 261 15 US-10-245-871-358 Sequence 358, App  
39 40 80.0 261 16 US-10-378-449-2 Sequence 2, Appl  
40 40 80.0 262 15 US-10-298-965-14 Sequence 14, Appl  
41 40 80.0 375 9 US-09-755-100-11 Sequence 11, Appl  
42 40 80.0 375 15 US-10-298-965-11 Sequence 11, Appl  
43 40 80.0 1079 9 US-09-822-827-947 Sequence 947, App  
44 40 80.0 1079 9 US-09-895-793-947 Sequence 947, App  
45 38 76.0 9 9 US-09-796-294-46 Sequence 46, Appl

ALIGNMENTS

RESULT 1  
US-09-918-243-80  
; Sequence 80, Application US/09918243  
; Patent No. US20020142317A1  
; GENERAL INFORMATION:  
; APPLICANT: O'Brien, Timothy J.  
; APPLICANT: Cannon, Martin J.  
; APPLICANT: Santin, Alessandro  
; TITLE OF INVENTION: Methods for the early diagnosis of ovarian cancer  
; FILE REFERENCE: D6223CIP/C/D/CIP  
; CURRENT APPLICATION NUMBER: US/09/918,243  
; CURRENT FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US  
; PRIOR FILING DATE: 2001-07-13  
; NUMBER OF SEQ ID NOS: 136  
; SEQ ID NO 60  
; LENGTH: 9  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CHAIN  
; OTHER INFORMATION: Residues 207-215 of the SCCE protein  
US-09-918-243-80

Query Match 100.0%; Score 50; DB 9; Length 9;  
Best Local Similarity 100.0%; Pred. No. 7.2e+05;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9  
DB 1 GPLVCRGTL 9

RESULT 2  
US-09-905-083-80  
; Sequence 80, Application US/09905083  
; Patent No. US20020146708A1  
; GENERAL INFORMATION:

APPLICANT: O'Brien, Timothy J.  
TITLE OF INVENTION: Compositions and Methods for the Early Diagnosis of  
TITLE OF INVENTION: Ovarian Cancer  
FILE REFERENCE: D6223CIP/C/D1V  
CURRENT APPLICATION NUMBER: US/09/905,083  
CURRENT FILING DATE: 2001-07-13  
PRIOR APPLICATION NUMBER: US 09/502,600  
PRIOR FILING DATE: 2000-02-11  
NUMBER OF SEQ ID NOS: 136  
SEQ ID NO 80  
LENGTH: 9  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CHAIN  
OTHER INFORMATION: Residues 207-215 of the SCCE protein

3-09-905-083-80

Query Match 100.0%; Score 50; DB 9; Length 9;  
Best Local Similarity 100.0%; Pred. No. 7.2e+05;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

y 1 GPLVCRGTL 9  
b 1 GPLVCRGTL 9

## RESULT 3

S-09-888-615-98  
Sequence 98, Application US/09888615  
Patent No. US20020064856A1  
GENERAL INFORMATION:

APPLICANT: FLOWMAN, GREGORY  
APPLICANT: WHYTE, DAVID  
APPLICANT: CAENEPEEL, SEAN  
APPLICANT: CHARYCZAK, GLEN  
APPLICANT: MANNING, GERARD  
APPLICANT: SUDARSANAM, SUCHA

TITLE OF INVENTION: NOVEL PROTEASES

FILE REFERENCE: 038602/1214  
CURRENT APPLICATION NUMBER: US/09/888,615

CURRENT FILING DATE: 2001-06-26

PRIOR APPLICATION NUMBER: 60/214,047

PRIOR FILING DATE: 2000-06-26

NUMBER OF SEQ ID NOS: 150

SOFTWARE: Patentin Ver. 2.1

SEQ ID NO 98

LENGTH: 253

TYPE: PRT

ORGANISM: Homo sapiens

S-09-888-615-98

Query Match 100.0%; Score 50; DB 9; Length 253;  
Best Local Similarity 100.0%; Pred. No. 0.54;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

y 1 GPLVCRGTL 9  
b 207 GPLVCRGTL 215

## RESULT 4

JS-09-764-762-3  
Sequence 3, Application US/09764762  
Patent No. US20020068341A1  
GENERAL INFORMATION:

APPLICANT: Hillman, Jennifer L.

Lal, Preeti

TITLE OF INVENTION: NOVEL KALLIKREIN

NUMBER OF SEQUENCES: 5

CORRESPONDENCE ADDRESS:

ADDRESSEE: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Drive

CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/764,762  
FILING DATE: 16-Jan-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/210,084  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0252 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 253 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 532504  
SEQUENCE DESCRIPTION: SEQ ID NO: 3:

US-09-764-762-3

Query Match 100.0%; Score 50; DB 9; Length 253;  
Best Local Similarity 100.0%; Pred. No. 0.54;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

y 1 GPLVCRGTL 9  
b 207 GPLVCRGTL 215

## RESULT 5

US-10-264-283-90  
Sequence 90, Application US/10264283  
Publication No. US2003014494A1  
GENERAL INFORMATION:

APPLICANT: Algate, Paul A.

APPLICANT: Mannion, Jane

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY

FILE REFERENCE: 210121.590

CURRENT APPLICATION NUMBER: US/10/264,283

CURRENT FILING DATE: 2002-10-02

NUMBER OF SEQ ID NOS: 111

SOFTWARE: Corixa Invention Disclosure Database

SEQ ID NO 90

LENGTH: 253

TYPE: PRT

ORGANISM: Homo sapiens

US-10-264-283-90

Query Match 100.0%; Score 50; DB 14; Length 253;  
Best Local Similarity 100.0%; Pred. No. 0.54;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

y 1 GPLVCRGTL 9  
b 207 GPLVCRGTL 215

```

; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: US 60/315,287
; PRIOR FILING DATE: 2001-08-27
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/372,246
; PRIOR FILING DATE: 2001-04-12
; NUMBER OF SEQ ID NOS: 163
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 48
; LENGTH: 253
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-173-999-48

Query Match      100.0%; Score 50; DB 15; Length 253;
Best Local Similarity 100.0%; Pred. No. 0.54;
Matches      9; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

QY      1      GPLVCRGTL 9
         |||||
DB      207      GPLVCRGTL 215

RESULT 8
US-09-764-898-284
; Sequence 284, Application US/09764898
; Patent No. US20020090673A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PJZ01
; CURRENT APPLICATION NUMBER: US/09/764,898
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 311
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 284
; LENGTH: 247
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (145)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (147)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (148)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (185)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (201)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-898-284

Query Match      90.0%; Score 45; DB 9; Length 247;
Best Local Similarity 88.9%; Pred. No. 3.9;
Matches      8; Conservative      0; Mismatches      1; Indels      0; Gaps      0;

QY      1      GPLVCRGTL 9
         |||||
DB      203      GPLVCRGTL 211

RESULT 9
US-09-764-847-742
; Sequence 742, Application US/09764847
; Patent No. US20020132767A1
; GENERAL INFORMATION:

```

APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
FILE REFERENCE: PC009  
CURRENT APPLICATION NUMBER: US/09/764,847  
CURRENT FILING DATE: 2001-01-17  
Prior application data removed - consult PALM or file wrapper  
NUMBER OF SEQ ID NOS: 2003  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 742  
LENGTH: 247  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (145)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (147)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (148)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (185)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (201)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
3-09-764-847-742

Query Match 90.0%; Score 45; DB 9; Length 247;  
Best Local Similarity 88.9%; Pred. No. 3.9;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

y 1 GPLVCRGTL 9  
b 203 GPLVCRGTL 211

ESULT 10  
S-10-092-154-742  
Sequence 742, Application US/10092154  
Publication No US20030054375A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
FILE REFERENCE: PC009C1  
CURRENT APPLICATION NUMBER: US/10/092,154  
CURRENT FILING DATE: 2002-03-07  
NUMBER OF SEQ ID NOS: 2003  
Prior application data removed - See File Wrapper or Palm  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 742  
LENGTH: 247  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (145)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: misc feature  
LOCATION: (147)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: misc feature  
LOCATION: (148)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: misc feature  
LOCATION: (185)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: misc feature  
LOCATION: (201)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
JS-10-092-154-742

Query Match 90.0%; Score 45; DB 14; Length 247;  
Best Local Similarity 88.9%; Pred. No. 3.9;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9  
DB 203 GPLVCRGTL 211

RESULT 11  
US-10-055-569A-69  
; Sequence 69, Application US/10055569A  
; Publication NO. US20040024181A1  
; GENERAL INFORMATION:  
; APPLICANT: Gangolli, Esha A  
; APPLICANT: Spyttek, Kimberly A  
; APPLICANT: Gilbert, Jennifer  
; APPLICANT: Casman, Stacie  
; APPLICANT: Bialock, Angela  
; APPLICANT: Li, Li  
; APPLICANT: Vernet, Corine  
; APPLICANT: Shenoy, Suresh  
; APPLICANT: Mishra, Vishnu S  
; APPLICANT: Furtak, Katarzyna  
; APPLICANT: Gerlach, Valerie L  
; APPLICANT: Edinger, Shlomit  
; APPLICANT: Malyanker, Uriel  
; APPLICANT: Stone, David  
; APPLICANT: Millet, Isabelle  
; APPLICANT: Smithson, Glenda  
; APPLICANT: Gunther, Erik  
; APPLICANT: Ellerman, Karen  
; APPLICANT: Padigaru, Muralidhara  
; APPLICANT: Taupier Jr., Raymond J  
; APPLICANT: Anderson, David W  
; FILE REFERENCE: 21402-191  
; CURRENT APPLICATION NUMBER: US/10/055,569A  
; CURRENT FILING DATE: 2001-10-26  
; PRIOR APPLICATION NUMBER: 60/243,642  
; PRIOR FILING DATE: 2000-10-26  
; PRIOR APPLICATION NUMBER: 60/243,320  
; PRIOR FILING DATE: 2000-10-26  
; PRIOR APPLICATION NUMBER: 60/243,592  
; PRIOR FILING DATE: 2000-10-26  
; PRIOR APPLICATION NUMBER: 60/243,681  
; PRIOR FILING DATE: 2000-10-27  
; PRIOR APPLICATION NUMBER: 60/243,863  
; PRIOR FILING DATE: 2000-10-27  
; PRIOR APPLICATION NUMBER: 60/244,443  
; PRIOR FILING DATE: 2000-10-31  
; PRIOR APPLICATION NUMBER: 60/245,029  
; PRIOR FILING DATE: 2000-11-01  
; PRIOR APPLICATION NUMBER: 60/244,995  
; PRIOR FILING DATE: 2000-11-01  
; PRIOR APPLICATION NUMBER: 60/245,293  
; PRIOR FILING DATE: 2000-11-02  
; PRIOR APPLICATION NUMBER: 60/245,315  
; PRIOR FILING DATE: 2000-11-02  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 137  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 69  
; LENGTH: 250  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-055-569A-69

Query Match 90.0%; Score 45; DB 16; Length 250;  
Best Local Similarity 88.9%; Pred. No. 4;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 GPLVCRGTL 9  
|||||  
206 GPLVNGTL 214

RESULT 12  
3-09-764-898-212  
Sequence 212, Application US/09/764,898

Patent No. US20020090673A1  
GENERAL INFORMATION:

APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

FILE REFERENCE: P0201  
CURRENT APPLICATION NUMBER: US/09/764,898

Prior application data removed - consult PALM or file wrapper  
NUMBER OF SEQ ID NOS: 311

SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 212

LENGTH: 251  
TYPE: PRT

ORGANISM: Homo sapiens  
S-09-764-898-212

Query Match 90.0%; Score 45; DB 9; Length 251;  
Best Local Similarity 89.9%; Pred. No. 4;

Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Y 1 GPLVCRGTL 9  
|||||  
b 207 GPLVNGTL 215

RESULT 13

S-10-055-569A-68

Sequence 68, Application US/10055569A  
Publication No. US20040024181A1

GENERAL INFORMATION:

APPLICANT: Gangolli, Esha A  
APPLICANT: Spytek, Kimberly A

APPLICANT: Gilbert, Jennifer  
APPLICANT: Casman, Stacie

APPLICANT: Blalock, Angela  
APPLICANT: Li, Li

APPLICANT: Vernet, Corine  
APPLICANT: Shenoy, Suresh

APPLICANT: Mishra, Vishnu S  
APPLICANT: Furtak, Katarzyna

APPLICANT: Gerlach, Valerie L  
APPLICANT: Edinger, Shlomit

APPLICANT: Malyanker, Uriel  
APPLICANT: Stone, David

APPLICANT: Millet, Isabelle  
APPLICANT: Smithson, Glennda

APPLICANT: Ellerman, Karen  
APPLICANT: Padigaru, Muralidhara

APPLICANT: Taudier Jr., Raymond J  
APPLICANT: Anderson, David W

TITLE OF INVENTION: No. US20040024181A1 Human Proteins, Polynucleotides Encoding Th  
FILE REFERENCE: 21402-191

CURRENT APPLICATION NUMBER: US/10/055,569A  
CURRENT FILING DATE: 2001-10-26

Prior application data removed - consult PALM or file wrapper  
NUMBER OF SEQ ID NOS: 60/243,642

Prior application data removed - consult PALM or file wrapper  
NUMBER OF SEQ ID NOS: 60/243,320

Prior application data removed - consult PALM or file wrapper  
NUMBER OF SEQ ID NOS: 60/243,592

Prior application data removed - consult PALM or file wrapper  
NUMBER OF SEQ ID NOS: 60/243,691

Prior application data removed - consult PALM or file wrapper  
NUMBER OF SEQ ID NOS: 60/243,691

1 PRIOR APPLICATION NUMBER: 60/243,863  
2 PRIOR FILING DATE: 2000-10-27  
3 PRIOR APPLICATION NUMBER: 60/244,443  
4 PRIOR FILING DATE: 2000-10-31  
5 PRIOR APPLICATION NUMBER: 60/245,029  
6 PRIOR FILING DATE: 2000-11-01  
7 PRIOR APPLICATION NUMBER: 60/244,995  
8 PRIOR FILING DATE: 2000-11-01  
9 PRIOR APPLICATION NUMBER: 60/245,293  
10 PRIOR FILING DATE: 2000-11-02  
11 PRIOR APPLICATION NUMBER: 60/245,315  
12 PRIOR FILING DATE: 2000-11-02  
13 Remaining Prior Application data removed - See File Wrapper or PALM.  
14 NUMBER OF SEQ ID NOS: 137  
15 SOFTWARE: Patentin Ver. 2.1  
16 SEQ ID NO 68  
17 LENGTH: 249  
18 TYPE: PRT  
19 ORGANISM: Mus musculus  
20 US-10-055-569A-68

Query Match 82.0%; Score 41; DB 16; Length 249;  
Best Local Similarity 77.8%; Pred. No. 20;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GPLVCRGTL 9  
|||||  
Db 204 GPLVNGSL 212

RESULT 14

US-10-055-569A-67

Sequence 67, Application US/10055569A  
Publication No. US20040024181A1

GENERAL INFORMATION:

APPLICANT: Gangolli, Esha A  
APPLICANT: Spytek, Kimberly A

APPLICANT: Gilbert, Jennifer  
APPLICANT: Casman, Stacie

APPLICANT: Blalock, Angela  
APPLICANT: Li, Li

APPLICANT: Vernet, Corine  
APPLICANT: Shenoy, Suresh

APPLICANT: Mishra, Vishnu S  
APPLICANT: Furtak, Katarzyna

APPLICANT: Gerlach, Valerie L  
APPLICANT: Edinger, Shlomit

APPLICANT: Malyanker, Uriel  
APPLICANT: Stone, David

APPLICANT: Millet, Isabelle  
APPLICANT: Smithson, Glennda

APPLICANT: Ellerman, Karen  
APPLICANT: Padigaru, Muralidhara

APPLICANT: Taudier Jr., Raymond J  
APPLICANT: Anderson, David W

TITLE OF INVENTION: No. US20040024181A1 Human Proteins, Polynucleotides Encoding Th  
FILE REFERENCE: 21402-191

CURRENT APPLICATION NUMBER: US/10/055,569A  
CURRENT FILING DATE: 2001-10-26

Prior application data removed - consult PALM or file wrapper  
NUMBER OF SEQ ID NOS: 60/243,642

Prior application data removed - consult PALM or file wrapper  
NUMBER OF SEQ ID NOS: 60/243,320

Prior application data removed - consult PALM or file wrapper  
NUMBER OF SEQ ID NOS: 60/243,592

Prior application data removed - consult PALM or file wrapper  
NUMBER OF SEQ ID NOS: 60/243,681

Prior application data removed - consult PALM or file wrapper  
NUMBER OF SEQ ID NOS: 60/243,863

Prior application data removed - consult PALM or file wrapper  
NUMBER OF SEQ ID NOS: 60/244,443

Prior application data removed - consult PALM or file wrapper  
NUMBER OF SEQ ID NOS: 2000-10-31



PRIOR APPLICATION NUMBER: 60/245,029  
PRIOR FILING DATE: 2000-11-01  
PRIOR APPLICATION NUMBER: 60/244,995  
PRIOR FILING DATE: 2000-11-01  
PRIOR APPLICATION NUMBER: 60/245,293  
PRIOR FILING DATE: 2000-11-02  
PRIOR APPLICATION NUMBER: 60/245,315  
PRIOR FILING DATE: 2000-11-02  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 137  
SOFTWARE: PatentIn ver. 2.1  
SEQ ID NO 67  
LENGTH: 276  
TYPE: PRT  
ORGANISM: Mus musculus  
-10-055-569A-67

Query Match 82.0%; Score 41; DB 16; Length 276;  
Best Local Similarity 77.8%; Pred. No. 22;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
/ 1 GPLVCRGTL 9  
||| | | | | | | | | |  
3 231 GPLVCGSL 239

RESULT 15  
S-09-860-739-5  
Sequence 5, Application US/09860739  
Publication No. US20030166036A1  
GENERAL INFORMATION:  
APPLICANT: Hybritech Incorporated  
APPLICANT: Mikolajczyk, Stephen  
APPLICANT: Rittenhouse, Harry  
TITLE OF INVENTION: A Protease and an Aminopeptidase Associated with Development of B  
TITLE OF INVENTION: Prostatic Hyperplasia (BPH)  
FILE REFERENCE: 2024-451  
CURRENT APPLICATION NUMBER: US/09/860,739  
CURRENT FILING DATE: 2001-05-18  
NUMBER OF SEQ ID NOS: 5  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 5  
LENGTH: 54  
TYPE: PRT  
ORGANISM: Homo sapiens  
S-09-860-739-5

Query Match 80.0%; Score 40; DB 10; Length 54;  
Best Local Similarity 77.8%; Pred. No. 6.8;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
/ 1 GPLVCRGTL 9  
||| | | | | | | | | |  
1b 8 GPLVCGVL 16

Search completed: March 1, 2004, 18:08:52  
Job time : 24.1111 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

protein - protein search, using sw model

March 1, 2004, 17:21:01 ; Search time 10.2222 Seconds  
(without alignments)  
84.690 Million cell updates/sec

US-09-905-083-80

50

1 GPLVCRGTL 9

BLOSUM62

Gapop 10.0 , Gapext 0.5

283366 seqs, 96191526 residues

total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

atabase :

PIR 78:\*

1: PIR1:\*

2: PIR2:\*

3: PIR3:\*

4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	50	100.0	253	2	serine proteinase
2	40	80.0	261	2	semogelase (EC 3
3	40	80.0	261	2	tissue kallikrein
4	40	80.0	262	2	probable aminoglyc
5	39	78.0	155	2	hypothetical prote
6	39	78.0	265	1	tissue kallikrein
7	38	76.0	246	1	complement factor
8	38	76.0	261	1	semogelase (EC 3
9	37	74.0	46	2	glandular kallikre
10	37	74.0	96	2	tissue kallikrein
11	37	74.0	104	2	tissue kallikrein
12	37	74.0	158	2	tissue kallikrein
13	37	74.0	186	2	tissue kallikrein
14	37	74.0	225	2	probable serine pr
15	37	74.0	239	2	tissue kallikrein
16	37	74.0	240	2	trypsin (EC 3.4.21
17	37	74.0	241	2	trypsin (EC 3.4.21
18	37	74.0	244	2	tissue kallikrein
19	37	74.0	250	2	tissue kallikrein
20	37	74.0	254	2	chymotrypsin-like
21	37	74.0	258	2	trypsin 2 - rat
22	37	74.0	259	1	tonin (EC 3.4.21-
23	37	74.0	259	2	tissue kallikrein
24	37	74.0	259	2	tissue kallikrein
25	37	74.0	260	2	neuropilin - mouse
26	37	74.0	261	1	tissue kallikrein
27	37	74.0	261	1	7S nerve growth fa
28	37	74.0	261	1	tissue kallikrein
29	37	74.0	261	2	tissue kallikrein

tissue kallikrein  
gamma-renin (EC 3.  
tissue kallikrein  
tissue kallikrein  
tissue kallikrein  
tissue kallikrein  
tissue kallikrein  
trypsin (EC 3.4.21  
trypsin (EC 3.4.21  
trypsin (EC 3.4.21  
snake venom factor

ALIGNMENTS

RESULT 1

A53968

serine proteinase SCCE precursor - human

N/Alternate names: stratum corneum chymotryptic enzyme

C/Species: Homo sapiens (man)

C/Date: 07-Jul-1995 #sequence\_revision 07-Jul-1995 #text\_change 22-Jun-1999

C/Accession: A53968

R/Hansson, L.; Stroemqvist, M.; Baeckman, A.; Wallbrandt, P.; Carlstein, A.; Egelrud, T.

J. Biol. Chem. 269, 19420-19426, 1994

A/Title: Cloning, expression, and characterization of stratum corneum chymotryptic enzym

A/Reference number: A53968; PMID:94308225; PMID:8034709

A/Accession: A53968

A/Status: Preliminary

A/Molecule type: mRNA

A/Residues: 1-253 <HAN>

A/Cross-references: GB:I33404; NID:9521214; PIDN:AAC37551.1; PID:G532504

C/Genetics:

A/Gene: GDB:PRSS6; SCCE

A/Cross-references: GDB:377730

A/Map position: 7q35-7q35

C/Superfamily: trypsin; trypsin homology

F:30-245/Domain: trypsin homology <TRY>

Query Match 100.0%; Score 50; DB 2; Length 253;

Best Local Similarity 100.0%; Pred. No. 0.085;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GPLVCRGTL 9

Db 207 GPLVCRGTL 215

RESULT 2

A32297

semogelase (EC 3.4.21.77) precursor [validated] - human

N/Alternate names: gamma-semioprostein; P-30 antigen; prostate-specific antigen; prosta

C/Species: Homo sapiens (man)

C/Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 08-Dec-2000

C/Accession: A32297; A32423; S03604; S05468; S05467; A32546; S02239; A26757; C31567; A31:

R/Riegman, P.H.J.; Vilestra, R.J.; van der Korp, J.A.G.M.; Romijn, J.C.; Trapman, J.

Biochem. Biophys. Res. Commun. 159, 95-102, 1989

A/Title: Characterization of the prostate-specific antigen gene: a novel human kallikrei

A/Reference number: A32297; PMID:89165891; PMID:2466464

A/Accession: A32297

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-261 <RIE>

A/Cross-references: GB:M24543

R/Lundwall, A.

Biochem. Biophys. Res. Commun. 161, 1151-1159, 1989

A/Title: Characterization of the gene for prostate-specific antigen, a human glandular k

A/Reference number: A32423; PMID:89302090; PMID:2472789

Accession: A32423  
 Status: Preliminary  
 Molecule type: DNA  
 Residues: 1-261 <LUN>  
 Cross-references: GB:M27274; NID:G190552; PIDN:AAA60192.1; PID:G190553  
 Digby, M.; Zhang, X.Y.; Richards, R.I.  
 Proc. Natl. Acad. Sci. U.S.A. 83, 3166-3170, 1986  
 Title: Human prostate specific antigen (PSA) gene: structure and linkage to the kallikrein gene  
 Reference number: S03604; MUID:89193632; PMID:2467258  
 Accession: S03604  
 Molecule type: DNA  
 Residues: 1-261 <DIG>  
 Cross-references: EMBL:X13940  
 Klobbeck, H.G.; Combratio, G.; Schulz, P.; Arbusow, V.; Fittler, F.  
 Submitted to the EMBL Data Library, May 1989  
 Reference number: S05468  
 Accession: S05468  
 Molecule type: DNA  
 Residues: 1-261 <KL1>  
 Cross-references: EMBL:X14810; NID:G35732; PIDN:CAA32915.1; PID:G296671  
 Klobbeck, H.G.; Combratio, G.; Schulz, P.; Arbusow, V.; Fittler, F.  
 Nucleic Acids Res. 17, 1981, 1989  
 Title: Genomic sequence of human prostate specific antigen (PSA).  
 Reference number: S05467; MUID:89282407; PMID:2471958  
 Accession: S05467  
 Molecule type: DNA  
 Residues: 1-29 <KL2>  
 Cross-references: EMBL:X14810  
 Henttu, P.; Viikko, P.  
 Biochem. Biophys. Res. Commun. 160, 903-910, 1989  
 Title: cDNA coding for the entire human prostate specific antigen shows high homology to the kallikrein gene  
 Reference number: A32546; MUID:89246551; PMID:2470373  
 Accession: A32546  
 Molecule type: mRNA  
 Residues: 1-72, 'T', '74-85', 'I', '87-174', 'P', '176-183', 'Q', '185-259', 'D', '261' <HEN>  
 Cross-references: GB:M26663  
 Schulz, P.; Stucka, R.; Reidmann, H.; Combratio, G.; Klobbeck, H.G.; Fittler, F.  
 Nucleic Acids Res. 16, 5225, 1988  
 Title: Sequence of a cDNA clone encompassing the complete mature human prostate specific antigen (PSA) gene  
 Reference number: S02239; MUID:88289366; PMID:2456523  
 Accession: S02239  
 Molecule type: mRNA  
 Residues: 1-63, 'T', '65-135', 'M', '137-261' <SCH>  
 Cross-references: EMBL:X07730  
 Lundwall, A.; Lilja, H.  
 FEBS Lett. 214, 317-322, 1987  
 Title: Molecular cloning of human prostate specific antigen cDNA.  
 Reference number: A26757; MUID:87190978; PMID:2436946  
 Accession: A26757  
 Molecule type: mRNA  
 Residues: 5-261 <LU2>  
 Riegman, P.H.J.; Klaassen, P.; van der Korput, J.A.G.M.; Romijn, J.C.; Trapman, J.  
 Biochem. Biophys. Res. Commun. 155, 181-188, 1988  
 Title: Molecular cloning and characterization of novel prostate antigen cDNA's.  
 Reference number: A90144; MUID:88326297; PMID:2458104  
 Accession: C31567  
 Molecule type: mRNA  
 Residues: 5-261 <R12>  
 Cross-references: GB:M21895; NID:G189523; PIDN:AAA59995.1; PID:G189524  
 Note: the authors translated the codon GGC for residue 28 as Arg and TGG for residue 29  
 Accession: A31567  
 Molecule type: mRNA  
 Residues: 5-210, 'WVILITELTPRA', '223', 'PMVLHGSIV', '233', 'WRGGV' <R13>  
 Cross-references: GB:M21896; NID:G189525; PIDN:AAA59996.1; PID:G189526  
 Note: the authors translated the codon GGC for residue 28 as Arg and TGG for residue 29  
 Accession: B31567  
 Molecule type: mRNA  
 Residues: 5-164, 'CTPGPDGAAGSPDAWV' <R14>  
 Cross-references: GB:M21897; NID:G189529; PIDN:AAA59997.1; PID:G189530  
 Note: the authors translated the codon GGC for residue 28 as Arg, TGG for residue 29  
 Schaller, J.; Akiyama, K.; Tsuda, R.; Hara, M.; Marti, T.; Ricki, E.E.  
 Eur. J. Biochem. 170, 111-120, 1987  
 Title: Isolation, characterization and amino-acid sequence of gamma-seminoprotein, a

A;Reference number: S00232; MUID:88082806; PMID:3691515  
 A;Accession: S00232  
 A;Molecule type: Protein  
 A;Residues: 25-261 <SC2>  
 R;Watt, K.W.K.; Lee, P.J.; M'Timkulu, T.; Chan, W.P.; Looor, R.  
 Proc. Natl. Acad. Sci. U.S.A. 83, 3166-3170, 1986  
 A;Title: Human prostate-specific antigen: structural and functional similarity with serin  
 A;Reference number: A23937; MUID:86205857; PMID:2422647  
 A;Accession: A23937  
 A;Molecule type: Protein  
 A;Residues: 25-93, 'T', '95-164', 'HL', '166', 'YDQM', '169-174', 'Q', '176-261' <WAT>  
 R;Moreno, J.M.  
 Submitted to the EMBL Data Library, November 1994  
 A;Reference number: G07735  
 A;Accession: G01551  
 A;Status: translated from GB/EMBL/DBDJ  
 A;Molecule type: mRNA  
 A;Residues: 1-261 <MOR>  
 A;Cross-references: EMBL:U17040; NID:G595945; PIDN:AAA56764.1; PID:G595946  
 R;Monne, M.; Croce, C.M.; Yu, H.; Diamandis, E.P.  
 Cancer Res. 54, 6344-6347, 1994  
 A;Title: Molecular characterization of prostate-specific antigen messenger RNA expressed  
 A;Reference number: I52712; MUID:95079406; PMID:7527295  
 A;Accession: I52712  
 A;Status: preliminary; translated from GB/EMBL/DBDJ  
 A;Molecule type: mRNA  
 A;Residues: 25-30, 'X', '32-49' <CHR>  
 A;Cross-references: GB:S75755; NID:G861469; PIDN:AAD14185.1; PID:G4261885  
 R;Christenson, A.; Lilja, H.  
 Eur. J. Biochem. 220, 45-53, 1994  
 A;Title: Complex formation between protein C inhibitor and prostate-specific antigen in  
 A;Reference number: S41212; MUID:94164172; PMID:7509746  
 A;Accession: S41212  
 A;Molecule type: protein  
 A;Residues: 25-30, 'X', '32-49' <CHR>  
 C;Comment: This enzyme preferentially cleaves after tyrosine residues.  
 C;Genetics:  
 A;Gene: GBS:APS; PSA  
 A;Cross-references: GDB:119695; OMIM:176820  
 A;Map position: 19q13.3-19q13.3  
 A;Intons: 16/1; 69/2; 165/1; 210/3  
 C;Superfamily: trypsin; trypsin homology  
 C;Keywords: glycoprotein; hydrolase; prostate; serine proteinase  
 F;1-17/Domain: signal sequence #status predicted <SIG>  
 F;18-24/Domain: propeptide #status predicted <PRO>  
 F;25-261/Product: semogelase #status experimental <MAT>  
 F;25-253/Domain: trypsin homology <TRY>  
 F;65,120,213/Active site: His, Asp, Ser #status predicted

Query Match 80.0%; Score 40; DB 1; Length 261;  
 Best Local Similarity 77.8%; Pred. No. 6.1;  
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GPLVCRGTL 9  
 Db 215 GPLVCGVL 223

RESULT 3  
 A29586  
 tissue kallikrein (EC 3.4.21.35) hGK-1 precursor - human  
 N;Alternate names: glandular kallikrein  
 C;Species: Homo sapiens (man)  
 C;Date: 08-Mar-1989 #sequence\_revision 08-Mar-1989 #text\_change 22-Jun-1999  
 R;Accession: A29586  
 R;Schedlich, L.J.; Bennetts, B.H.; Morris, B.J.  
 DNA 6, 429-437, 1987  
 A;Title: Primary structure of a human glandular kallikrein gene.  
 A;Reference number: A29586; MUID:88054467; PMID:2824146  
 A;Accession: A29586  
 A;Molecule type: DNA  
 A;Residues: 1-261 <SCH>  
 A;Cross-references: GB:M18157; NID:G186640; PIDN:AAA74454.1; PID:G386842



Query Match 78.0%; Score 39; DB 1; Length 265;  
 Best Local Similarity 66.7%; Pred. No. 9.4; Indels 0;  
 Matches 6; Conservative 1; Mismatches 2; Gaps 0;

1 GPLVCRGTL 9  
 |||||  
 219 GPLICNGVL 227

SULT 7  
 HU  
 Complement factor D (EC 3.4.21.46) precursor [validated] - human (fragment)  
 Alternate names: adipsin; C3 convertase activator  
 Species: Homo sapiens (man)  
 Date: 28-Aug-1995 #sequence\_revision 31-Dec-1992 #text\_change 08-Dec-2000  
 Accession: A40197; A00936; A00571; S66645  
 White, R.T.; Damm, D.; Hancock, N.; Rosen, B.S.; Lowell, B.B.; Usher, P.; Flier, J.S.;  
 Biol. Chem. 267, 9210-9213, 1992  
 Title: Human adipsin is identical to complement factor D and is expressed at high level  
 Reference number: A40197; MUID:92250520; PMID:1374368  
 Accession: A40197  
 Molecule type: mRNA  
 Residues: 1-246 <WHI>  
 Cross-references: GB:M84526  
 Niemann, M.A.; Shown, A.S.; Bennett, J.C.; Volanakis, J.E.  
 Biochemistry 23, 2482-2486, 1984  
 Title: Amino acid sequence of human D of the alternative complement pathway.  
 Reference number: A00936; MUID:95000441; PMID:6383466  
 Accession: A00936  
 Molecule type: protein  
 Residues: 19-44, 'G', 46-51, 'Q', 53-75, 'TH', 78, 'P', 80-83, 'XXXIIIIE', 90-172, 86-91, 185-235, '  
 Note: A few residues were assigned from the previously published sequence of Reid et al  
 Miyata, T.; Oda, O.; Inagi, R.; Sugiyama, S.; Miyama, A.; Maeda, K.; Nakashima, I.; Ya  
 ol. Immunol. 27, 637-644, 1990  
 Title: Molecular and functional identification and purification of complement component  
 Reference number: A60571; MUID:90370044; PMID:2395435  
 Accession: A60571  
 Molecule type: protein  
 Residues: 19-20, 'XX', 23-27, 'XX', 30-31, 'XX', 34, 'X', 36-40 <MIY>  
 Balke, N.; Holtkamp, U.; Hoerl, W.H.; Tschesche, H.  
 EBS Lett. 371, 300-302, 1995  
 Title: Inhibition of degranulation of human polymorphonuclear leukocytes by complement  
 Reference number: S66645; MUID:96013156; PMID:7556615  
 Accession: S66645  
 Status: preliminary  
 Molecule type: protein  
 Residues: 19-44, 'C', 46-48 <BAL>  
 Comment: Factor D cleaves factor B when the latter is complexed with factor C3b, activ  
 Genetics:  
 Gene: GDB:DF  
 Cross-references: GDB:132645; OMIM:134350  
 Map position: Xpter-Xqter  
 Superfamily: trypsin; trypsin homology  
 Keywords: complement alternative pathway; hydrolase; plasma; serine proteinase  
 1-18/Domain: signal sequence #status predicted <SIG>  
 19-246/Product: complement factor D (fragment) #status experimental <MAT>  
 19-241/Domain: trypsin homology <TRY>  
 44-60, 141-207, 172-188, 197-222/Dissulfide bonds: #status predicted  
 159,105,201/Active site: His, Asp, Ser #status predicted

Query Match 76.0%; Score 38; DB 1; Length 246;  
 Best Local Similarity 77.8%; Pred. No. 13;  
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

1 GPLVCRGTL 9  
 |||||  
 203 GPLVCGGVL 211

RESULT 8  
 S35711  
 semenogelase (EC 3.4.21.7) precursor - rhesus macaque  
 Alternate names: gamma-seminoprotein; P-30 antigen; prostate-specific antigen; prostat

C;Species: Macaca mulatta (rhesus macaque)  
 C;Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 19-May-2000  
 C;Accession: S35711; S34239  
 R;Gauthier, E.R.; Chapdelaine, P.; Tremblay, R.R.; Dube, J.Y.  
 Biochim. Biophys. Acta 1174, 207-210, 1993  
 A;Title: Characterization of rhesus monkey prostate specific antigen cDNA.  
 A;Reference number: S35711; MUID:93363642; PMID:7689340  
 A;Accession: S35711  
 A;Molecule type: mRNA  
 A;Residues: 1-261 <GAU>  
 A;Cross-references: EMBL:X73560; NID:9311843; PIDN:CAA51957.1; PID:9311844  
 C;Comment: This enzyme preferentially cleaves after tyrosine residues.  
 C;Superfamily: trypsin; trypsin homology  
 C;Keywords: glycoprotein; hydrolase; prostate; serine proteinase  
 F;1-17/Domain: signal sequence #status predicted <SIG>  
 F;18-24/Domain: propeptide #status predicted <PRO>  
 F;25-261/Product: semenogelase #status predicted <MAT>  
 F;25-253/Domain: trypsin homology <TRY>  
 F;65,120,213/Active site: His, Asp, Ser #status predicted

Query Match 76.0%; Score 38; DB 1; Length 261;  
 Best Local Similarity 77.8%; Pred. No. 14;  
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9  
 |||||  
 Db 215 GPLVCDGVL 223

RESULT 9  
 I49416  
 Glandular kallikrein - western wild mouse (fragment)  
 C;Species: Mus spretus (western wild mouse)  
 C;Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 22-Jun-1999  
 C;Accession: I49416  
 R;Ko, M.S.; Wang, X.; Horton, J.H.; Hagen, M.D.; Takahashi, N.; Maezaki, Y.; Nadeau, J.H  
 Mamm. Genome 5, 349-355, 1994  
 A;Title: Genetic mapping of 40 cDNA clones on the mouse genome by PCR.  
 A;Reference number: I48934; MUID:94315082; PMID:8043949  
 A;Accession: I49416  
 A;Status: preliminary; translated from GB/EMBL/DBD  
 A;Molecule type: DNA  
 A;Residues: 1-46 <RES>  
 A;Cross-references: EMBL:U05716; NID:g497047; PIDN:AAB60472.1; PID:g497048  
 C;Superfamily: trypsin; trypsin homology

Query Match 74.0%; Score 37; DB 2; Length 46;  
 Best Local Similarity 66.7%; Pred. No. 4.5;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9  
 |||||  
 Db 1 GPLICDGVL 9

RESULT 10  
 A05308  
 tissue kallikrein (EC 3.4.21.35), submandibular MGK-2 - mouse (fragment)  
 Alternate names: Glandular kallikrein  
 C;Species: Mus musculus (house mouse)  
 C;Date: 05-Jun-1987 #sequence\_revision 05-Jun-1987 #text\_change 22-Jun-1999  
 C;Accession: A05308  
 R;Mason, A.J.; Evans, B.A.; Cox, D.R.; Shine, J.; Richards, R.I.  
 Nature 303, 300-307, 1983  
 A;Title: Structure of mouse kallikrein gene family suggests a role in specific processing  
 A;Reference number: A00941; MUID:83219214; PMID:6602295  
 A;Accession: A05308  
 A;Molecule type: DNA  
 A;Residues: 1-96 <MAS>  
 A;Cross-references: GB:V00829; NID:952775; PIDN:CAA24212.1; PID:9388523  
 A;Experimental source: Quakenbush inbred strain  
 A;Note: This sequence has been translated from two exons (11-147 and 522-674) located up  
 C;Comment: Glandular kallikreins cleave Met-Lys and Arg-Ser bonds in kininogen to release

```

;Genetics:
;Gene: mck-2
;Map position: 7
;Introns: 45/3
;Superfamily: trypsin; trypsin homology
;Keywords: glycoprotein; hydrolase; saliva; serine proteinase; submandibular gland
;1-88/Domain: trypsin homology (fragment) <TRY>

Query Match 74.0%; Score 37; DB 2; Length 96;
Best Local Similarity 66.7%; Pred. No. 8.8;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

NY 1 GPLVCRGTL 9
|||:|
50 GPLICDGV 58

RESULT 11
;Issue kallikrein-related proteinase (EC 3.4.21.-) k10 - rat (fragments)
;Species: Rattus norvegicus (Norway rat)
;Date: 18-Feb-1994 #sequence_revision 19-Jul-1996 #text_change 17-Oct-1997
;Accession: S15395; S15448
;Gutman, N.; Elmoujahed, A.; Brillard, M.; Monegier du Sorbier, B.; Gauthier, F.
;Sur. J. Biochem. 197, 425-429, 1991
;Title: Microheterogeneity of rat submaxillary gland kallikrein k10, a member of the ka
;Reference number: S15395; MUID:91224135; PMID:2026164
;Accession: S15395
;Molecule type: protein
;Residues: 1-23,24-51,52-104 <GUT>
;Experimental source: submaxillary gland
;Superfamily: trypsin; trypsin homology
;Keywords: hydrolase; serine proteinase
;F1-104/Domain: trypsin homology (fragments) <TRY>
;F1-23/Product: tissue kallikrein-related protein k10 light chain (fragment) #status exp
;F24-51/Product: tissue kallikrein-related protein k10 heavy chain (fragment) #status exp
;F52-104/Product: tissue kallikrein-related protein k10 heavy chain (fragment) #status exp

Query Match 74.0%; Score 37; DB 2; Length 104;
Best Local Similarity 66.7%; Pred. No. 9.4;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
|||:|
70 GPLICDGV 78

RESULT 12
;Issue kallikrein (EC 3.4.21.35) S1, submaxillary - rat (fragment)
;Alternate names: glandular kallikrein; kininogenin
;Species: Rattus norvegicus (Norway rat)
;Date: 22-Jul-1987 #sequence_revision 22-Jul-1987 #text_change 22-Jun-1999
;Accession: B23863
;Ashley, P.L.; MacDonald, R.J.
;Biochemistry 24, 4512-4520, 1985
;Title: Kallikrein-related mRNAs of the rat submaxillary gland: nucleotide sequences of
;Reference number: A23863; MUID:86051477; PMID:2998455
;Accession: B23863
;Molecule type: mRNA
;Residues: 1-156 <ASH>
;Cross-references: GB:M11564; NID:G205031; PIDN:AAA41465.1; PID:G205032
;Superfamily: trypsin; trypsin homology
;Keywords: hydrolase; serine proteinase
;F1-148/Domain: trypsin homology (fragment) <TRY>

Query Match 74.0%; Score 37; DB 2; Length 156;
Best Local Similarity 66.7%; Pred. No. 14;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
|||:|
110 GPLICDGV 118

RESULT 13
;Issue kallikrein (EC 3.4.21.35) 3 - rat (fragment)
;Species: Rattus norvegicus (Norway rat)
;Date: 22-Nov-1989 #sequence_revision 31-Dec-1993 #text_change 22-Jun-1999
;Accession: B32340
;R.Shai, S.Y.; Woodley-Miller, C.; Chao, J.; Chao, L.
;Biochemistry 28, 5334-5343, 1989
;Title: Characterization of genes encoding rat tonin and a kallikrein-like serine prote
;Reference number: A32340; MUID:89375248; PMID:2550051
;Accession: B32340
;Status: preliminary
;Molecule type: DNA
;Residues: 1-188 <SHA>
;Cross-references: GB:M26534; NID:G206773; PIDN:AAA42080.1; PID:G206774; GB:J02860
;Superfamily: trypsin; trypsin homology
;Keywords: hydrolase; serine proteinase
;F1-180/Domain: trypsin homology (fragment) <TRY>

Query Match 74.0%; Score 37; DB 2; Length 188;
Best Local Similarity 66.7%; Pred. No. 16;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
|||:|
142 GPLICDGV 150

RESULT 14
;Probable serine proteinase (EC 3.4.21.-) ACO - human (fragment)
;Species: Homo sapiens (man)
;Date: 14-Apr-1998 #sequence_revision 08-May-1998 #text_change 19-May-2000
;Accession: S45356
;R.Dihanich, M.; Spiess, M.
;Biochim. Biophys. Acta 1218, 225-228, 1994
;Title: A novel serine proteinase-like sequence from human brain.
;Reference number: S45356; MUID:94289486; PMID:8018728
;Accession: S45356
;Molecule type: mRNA
;Residues: 1-225 <DIH>
;Cross-references: EMBL:X75363; NID:G407137; PIDN:CAA53145.1; PID:G940540
;Experimental source: Alzheimer's disease patient brain cortex
;Genetics:
;Gene: ACO
;Introns: 175/3
;Superfamily: trypsin; trypsin homology
;Keywords: hydrolase; serine proteinase

Query Match 74.0%; Score 37; DB 2; Length 225;
Best Local Similarity 77.8%; Pred. No. 19;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
|||:|
180 GPLVCGGIL 188

RESULT 15
;Issue kallikrein (EC 3.4.21.35), prostatic - guinea pig
;Alternate names: glandular kallikrein
;Species: Cavia porcellus (guinea pig)
;Date: 08-Mar-1989 #sequence_revision 08-Mar-1989 #text_change 01-Aug-1997
;Accession: A27207
;Dunbar, J.C.; Bradshaw, R.A.
;Biochemistry 26, 3471-3478, 1987
;Title: Amino acid sequence of guinea pig prostatic kallikrein.
;Reference number: A27207; MUID:86000549; PMID:3307909
;Accession: A27207
;Molecule type: protein

```

Residues: 1-239 <DUN>  
Note: 50-Tip was also found  
Superfamily: trypsin; trypsin homology  
Keywords: hydrolase; serine proteinase  
1-231/Domain; trypsin homology <TRY>  
41,96,191/Active site: His, Asp, Ser #status predicted

Query Match 74.0%; Score 37; DB 2; Length 239;  
Best Local Similarity 66.7%; Pred. No. 20;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

1 GPLVCRGTL 9  
|||:|  
193 GPLICDGV 201

Search completed: March 1, 2004, 17:36:31  
Job time : 11.3333 secs





cleaves insulin B chain at 6-Leu-[Cys-7, 16-Tyr]-[Leu-17, 25-Phe]-[Tyr-26, and 26-Tyr]-[Thr-27]. Could play a role in the activation of precursors to inflammatory cytokines.

-!- TISSUE SPECIFICITY: Abundantly expressed in the skin and is expressed by keratinocytes in the epidermis. Very low levels are also seen in the brain and kidney.

-!- SIMILARITY: Belongs to peptidase family S1. Kallikrein subfamily.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; L33404; AAC37551.1; -

EMBL; AF166330; AAD49718.1; -

EMBL; AF243527; AAG33360.1; -

EMBL; AF332583; AAKG9624.1; -

PIR; A53968; A53968.

HSSP; P00763; IDPO.

MEROPS; S01.300; -

Genew; HGNC:6368; KLK7.

MIM; 604438; -

GO; GO:0008236; P:serine-type peptidase activity; TAS.

GO; GO:0008544; P:epidermal differentiation; TAS.

InterPro; IPR009003; Cys\_Ser\_trypsin.

InterPro; IPR001254; Peptidase\_S1.

InterPro; IPR001314; Peptidase\_S1A.

Pfam; PF00089; trypsin; 1.

PRINTS; PR00722; CHYMOTRYPSIN.

SMART; SM00020; TRYP\_SPC; 1.

PROSITE; PS00240; TRYP\_SIN\_DOM; 1.

PROSITE; PS00134; TRYP\_SIN\_HIS; 1.

PROSITE; PS00135; TRYP\_SIN\_SER; 1.

Hydrolase; Serine protease; Zymogen; Glycoprotein; Signal.

SIGNAL 1 22

ACTIVATION PEPTIDE.

PROPEP 23 29

CHAIN 30 253

KALLIKREIN 7.

ACT\_SITE 70 70

CHARGE RELAY SYSTEM (BY SIMILARITY).

ACT\_SITE 112 112

CHARGE RELAY SYSTEM (BY SIMILARITY).

ACT\_SITE 205 205

CHARGE RELAY SYSTEM (BY SIMILARITY).

DISULFID 36 137

BY SIMILARITY.

DISULFID 55 71

BY SIMILARITY.

DISULFID ? 239

BY SIMILARITY.

DISULFID 144 211

BY SIMILARITY.

DISULFID 176 190

BY SIMILARITY.

DISULFID 201 226

BY SIMILARITY.

CARBOHYD 246 246

N-LINKED (GLCNAC...) (POTENTIAL).

SEQUENCE 253 AA; 27525 MW; 2D68B6B15A76A668 CRC64;

Query Match 100.0%; Score 50; DB 1; Length 253;

Best Local Similarity 100.0%; Pred. NO. 0.018;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9

Db 207 GPLVCRGTL 215

RESULT 2

KLK9 HUMAN

STANDARD; PRT; 250 AA.

OC

OX

RN

RP

RA

RT

RL

RN

RP

RA

RT

RL

RN

RP

RA

RT

RL

RN

RP

RA

RT

RL

RN

RP

RA

RT

RL

RN

RP

RA

RT

RL

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

NCBI\_TaxID=9606;

[1]

SEQUENCE FROM N.A.

MEDLINE=20118156; PubMed=10652563;

Yousef G.M., Luo L.-Y., Diamandis E.P.;

"Identification of novel human kallikrein-like genes on chromosome

19q13.3-q13.4.";

Anticancer Res. 19:2843-2852(1999).

[2]

SEQUENCE FROM N.A.

MEDLINE=20247258; PubMed=10783266;

Yousef G.M., Diamandis E.P.;

"The expanded human kallikrein gene family: locus characterization and

molecular cloning of a new member, KLK-L3.";

Genomics 65:184-194(2000).

[3]

SEQUENCE FROM N.A.

MEDLINE=20510030; PubMed=11054574;

Gan L., Lee I., Smith R., Argonza-Barrett R., Lei H., McCuaig J.,

Moss P., Paepker B., Wang K.;

"Sequencing and expression analysis of the serine protease gene

cluster located in chromosome 19q13 region.";

Gene 257:119-130(2000).

[4]

SEQUENCE FROM N.A.

Lamerdin J.E., McCready P.M., Skowronski E., Viswanathan V.,

Burkhardt-Schultz K., Gordon L., Dias J., Ramirez M., Stilwagen S.,

Phan H., Velasco N., Do L., Regala W., Terry A., Brower A., Barnes J.,

Danganan L., Erler A., Christensen M., Georgescu A., Avila J., Liu S.,

Angarise T., Frankheim M., Attix C., Amico-Keller G., Coefield J.,

Duarte S., Lucas S., Bruce R., Thomas P., Quan G., Kronmiller B.,

Arellano A., Sanders C., Ow D., Nolan M., Trong S., Kobayashi A.,

Olson A.S., Carrano A.V.;

"Sequence analysis of chromosome 19q13.4.";

Submitted (Oct-2000) to the EMBL/GenBank/DBJ databases.

-!- SUBCELLULAR LOCATION: Secreted (Probable).

-!- TISSUE SPECIFICITY: Skin, thymus, trachea, cerebellum and spinal

cord.

-!- SIMILARITY: Belongs to peptidase family S1. Kallikrein subfamily.

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

EMBL; AF135026; AAD26427.2; -

EMBL; AF243527; AAG33362.1; -

EMBL; AC011473; AAG23255.1; -

HSSP; P00763; IDPO.

MEROPS; S01.307; -

Genew; HGNC:6370; KLK9.

MIM; 605504; -

GO; GO:0005576; C:extracellular; NAS.

GO; GO:0004252; F:serine-type endopeptidase activity; NAS.

GO; GO:0005508; P:proteolysis and peptidolysis; NAS.

InterPro; IPR009003; Cys\_Ser\_trypsin.

InterPro; IPR001254; Peptidase\_S1.

InterPro; IPR001314; Peptidase\_S1A.

Pfam; PF00089; trypsin; 1.

PRINTS; PR00722; CHYMOTRYPSIN.

SMART; SM00020; TRYP\_SPC; 1.

PROSITE; PS00240; TRYP\_SIN\_DOM; 1.

PROSITE; PS00134; TRYP\_SIN\_HIS; 1.

PROSITE; PS00135; TRYP\_SIN\_SER; 1.

Hydrolase; Serine protease; Glycoprotein; Signal.

SIGNAL 1 15

POTENTIAL.

CHAIN 16 250

KALLIKREIN 9.

ACT\_SITE 63 63

CHARGE RELAY SYSTEM (BY SIMILARITY).

ACT\_SITE 111 111

CHARGE RELAY SYSTEM (BY SIMILARITY).

-----

```

1 ACT SITE 204 204 CHARGE RELAY SYSTEM (BY SIMILARITY).
2 DISULFID 29 164 BY SIMILARITY.
3 DISULFID 48 64 BY SIMILARITY.
4 DISULFID 136 238 BY SIMILARITY.
5 DISULFID 143 210 BY SIMILARITY.
6 DISULFID 175 189 BY SIMILARITY.
7 DISULFID 200 225 BY SIMILARITY.
8 CARBOHYD 131 131 N-LINKED (GLCNAC. . .) (POTENTIAL).
9 CARBOHYD 166 166 N-LINKED (GLCNAC. . .) (POTENTIAL).
10 CARBOHYD 211 211 N-LINKED (GLCNAC. . .) (POTENTIAL).
11 SEQUENCE 250 AA; 27512 MW; 27852459063988 CRC64;

Query Match 90.0%; Score 45; DB 1; Length 250;
Best Local Similarity 88.9%; Pred. No. 0.16;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Y 1 GPLVCRGTL 9
b 206 GPLVCRGTL 214

RESULT 3
L KE HUMAN STANDARD; PRT; 251 AA.
C QP0G3, 16-OCT-2001 (Rel. 40, Created)
T 16-OCT-2001 (Rel. 40, Last sequence update)
T 28-FEB-2003 (Rel. 41, Last annotation update)
E Kallikrein 14 precursor (EC 3.4.21.-) (Kallikrein-like protein 6)
DE (KLK-L6).
IN KLK14 OR KLK16.
AS Homo sapiens (Human).
XC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
XC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
IN [1]
IP SEQUENCE FROM N.A.
JA Yousef G.M., Diamandis E.P.;
UT "Molecular characterization, mapping, and tissue expression of KLK16,
UT a hormonally regulated kallikrein-like gene.";
UL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
UN [2]
IP SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
RX MEDLINE=21250997; PubMed=11352573;
XA Hooper J.D., Bui L.T., Rae F.K., Harvey T.J., Myers S.A.,
XA Ashworth L.K., Clements J.A.;
XT "Identification and characterization of KLK14, a novel kallikrein
XT serine protease gene located on human chromosome 19q13.4 and expressed
XT in prostate and skeletal muscle.";
RL Genomics 73:117-122(2001).
RN [3]
RP SEQUENCE FROM N.A.
RA Lamerdin J.E., McCready P.M., Skowronski E., Viswanathan V.,
RA Burkhardt-Schultz K., Gordon L., Dias J., Ramirez M., Stillwagen S.,
RA Phan H., Velasco N., Do L., Regalia M., Terry A., Brower A., Ganes J.,
RA Danganan L., Erler A., Christensen M., Georgescu A., Avila J., Liu S.,
RA Andreise T., Frankheim M., Attix C., Amico-Keller G., Coefield J.,
RA Duarte S., Lucas S., Bruce R., Thomas P., Quan G., Krommiller B.,
RA Arellano A., Sanders C., Ow D., Nolan M., Trong S., Kobayashi A.,
RA Olsen A.S., Carrano A.V.;
RT "Sequence analysis of chromosome 19q13.4.";
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP TISSUE SPECIFICITY.
RX MEDLINE=20545474; PubMed=10969073;
RA Harvey T.J., Hooper J.D., Myers S.A., Stephenson S.A., Ashworth L.K.,
RA Clements J.A.;
RT "Tissue-specific expression patterns and fine mapping of the human
RT kallikrein (KLK) locus on proximal 19q13.4.";
RL J. Biol. Chem. 275:37397-37406(2000).
CC -!- SUBCELLULAR LOCATION: Secreted (Probable).
CC -!- TISSUE SPECIFICITY: High expression in brain, bone marrow and
CC fetal liver. Also expressed in liver, pancreas, fetal spleen,

```

```

CC -!- SIMILARITY: Belongs to peptidase family S1. Kallikrein subfamily.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF161221; AAD50773.2; -
DR EMBL; AF283669; AAK48523.1; -
DR EMBL; AF283670; AAK48524.1; -
DR EMBL; AC011473; AAG23260.1; -
DR HSSP; P00763; LDPO.
DR MEROPS; S01.029; -
DR GO; GO:0005576; C:extracellular; NAS.
DR GO; GO:0004252; F:serine-type endopeptidase activity; NAS.
DR GO; GO:0006508; P:proteolysis and peptidolysis; NAS.
DR InterPro; IPR009003; Cys Ser trypsin.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PRO0722; CHYMOTRYPSIN.
DR SMART; SM00020; TRYPSIN_SPC; 1.
DR PROSITE; PS0240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Signal; Zymogen.
FT SIGNAL 1 18 POTENTIAL.
FT PROPEP 19 24 ACTIVATION PEPTIDE (POTENTIAL).
FT CHAIN 25 251 KALLIKREIN 14.
FT ACT_SITE 67 67 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 111 111 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 204 204 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 31 164 BY SIMILARITY.
FT DISULFID 52 68 BY SIMILARITY.
FT DISULFID 143 210 BY SIMILARITY.
FT DISULFID 175 189 BY SIMILARITY.
FT DISULFID 200 225 BY SIMILARITY.
SQ SEQUENCE 251 AA; 27452 MW; 9087953EAF7ED25 CRC64;

Query Match 88.0%; Score 44; DB 1; Length 251;
Best Local Similarity 88.9%; Pred. No. 0.26;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
DB 206 GPLVCRGTL 214

RESULT 4
VSP3 BOTJA
ID VSP3 BOTJA STANDARD; PRT; 258 AA.
AC Q9PTU8;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Venom serine proteinase A precursor (EC 3.4.21.-).
OS Bothrops jararaca (Jararaca).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidosauria; Squamata; Chordata; Serpentes; Colubroidea;
OC Viperidae; Crotalinae; Bothrops.
OC NCBI_Taxid=8724;
RN [1]
RP SEQUENCE FROM N.A.
RA Murayama N.;
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Thrombin-like snake venom serine protease.

```

-!- SUBCELLULAR LOCATION: Secreted.  
 -!- TISSUE SPECIFICITY: Expressed by the venom gland.  
 -!- SIMILARITY: Belongs to peptidase family S1. Snake venom subfamily.  
 -----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----

EMBL: AB031394; BAA89310.1; --  
 HSSP: P00763; IDPO.  
 MEROPS: S01.181; --  
 InterPro: IPR009003; Cys\_Ser\_trypsin.  
 InterPro: IPR001254; Peptidase\_S1.  
 InterPro: IPR001314; Peptidase\_S1A.  
 Pfam: PF00089; trypsin; 1.  
 PRINTS: PR00722; CHYMOTRYPSIN.  
 SMART: SM00020; TRYPSIN\_DOM; 1.  
 PROSITE: PS00240; TRYPSIN\_DOM; 1.  
 PROSITE: PS00134; TRYPSIN\_HIS; 1.  
 PROSITE: PS00135; TRYPSIN\_SER; 1.  
 HydroLase: Serine protease; Glycoprotein; Zymogen; Signal.  
 SIGNAL 1 18  
 PROPEP 19 24  
 CHAIN 25 258  
 ACT\_SITE 65 65  
 ACT\_SITE 110 110  
 ACT\_SITE 204 204  
 DISULFID 31 163  
 DISULFID 50 66  
 DISULFID 98 256  
 DISULFID 142 210  
 DISULFID 174 189  
 DISULFID 200 225  
 CARBOHYD 32 32  
 CARBOHYD 44 44  
 CARBOHYD 103 103  
 CARBOHYD 121 121  
 CARBOHYD 154 154  
 CARBOHYD 170 170  
 CARBOHYD 211 211  
 CARBOHYD 251 251  
 SEQUENCE 258 AA; 28058 MW; 034F9FDCB79EB64 CRC64;  
 Query Match 80.0%; Score 40; DB 1; Length 258;  
 Best Local Similarity 75.0%; Pred. No. 1.6;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Y 1 GPLVCRGT 8  
 | | | | |  
 b 206 GPLICNGT 213

RESULT 5  
 NRPN RAT STANDARD; PRT; 260 AA.  
 AC O88780;  
 DT 15-JUL-1999 (Rel. 38, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Neuropeptide precursor (EC 3.4.21.-) (NP) (Kallikrein 8) (Brain serine protease 1).  
 EN KLR8 OR PRS19 OR NRPN OR BSL1.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 DX NCBI\_TaxID=10116;  
 XN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Fischer; TISSUE=Brain;

RX MEDLINE=98389725; PubMed=9722524;  
 RA Davies B.J., Pickard B.S., Steel M., Morris R.G.M., Lathe R.;  
 RT "Serine proteases in rodent hippocampus.";   
 RL J. Biol. Chem. 273:23004-23011(1998).  
 CC -!- FUNCTION: Suggested to be involved in kindling epileptogenesis and hippocampal plasticity. Has a strong proteolytic activity against fibronectin (by similarity).  
 CC -!- CATALYTIC ACTIVITY: Preferential cleavage: Arg-, Lys-.  
 CC -!- SUBCELLULAR LOCATION: Secreted (by similarity).  
 CC -!- TISSUE SPECIFICITY: Restricted to hippocampus.  
 CC -!- SIMILARITY: Belongs to peptidase family S1. Kallikrein subfamily.  
 -----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 EMBL: AJ005641; CAA06643.1; --  
 HSSP: Q61395; INPM.  
 MEROPS: S01.244; --  
 InterPro: IPR009003; Cys\_Ser\_trypsin.  
 InterPro: IPR001254; Peptidase\_S1.  
 InterPro: IPR001314; Peptidase\_S1A.  
 Pfam: PF00089; trypsin; 1.  
 PRINTS: PR00722; CHYMOTRYPSIN.  
 SMART: SM00020; TRYPSIN\_DOM; 1.  
 PROSITE: PS00240; TRYPSIN\_DOM; 1.  
 PROSITE: PS00134; TRYPSIN\_HIS; 1.  
 PROSITE: PS00135; TRYPSIN\_SER; 1.  
 HydroLase: Serine protease; Glycoprotein; Zymogen; Signal.  
 SIGNAL 1 28  
 PROPEP 29 32  
 CHAIN 33 260  
 ACT\_SITE 73 73  
 ACT\_SITE 120 120  
 ACT\_SITE 212 212  
 DISULFID 59 173  
 DISULFID 59 74  
 DISULFID 145 246  
 DISULFID 152 218  
 DISULFID 184 198  
 DISULFID 208 233  
 CARBOHYD 110 110  
 SEQUENCE 260 AA; 28510 MW; 58DF4F0602A0B7F5 CRC64;  
 Query Match 80.0%; Score 40; DB 1; Length 260;  
 Best Local Similarity 77.8%; Pred. No. 1.6;  
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Y 1 GPLVCRGT 9  
 | | | | |  
 b 214 GPLVCGVL 222

RESULT 6  
 KLR2 HUMAN STANDARD; PRT; 261 AA.  
 ID KLR2 HUMAN  
 AC P20151; Q15946; Q9UJ29;  
 DT 01-FEB-1991 (Rel. 17, Created)  
 DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Glandular kallikrein 2 precursor (EC 3.4.21.35) (Tissue kallikrein) (Prostate) (HGK-1).  
 EN KLR2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 DX NCBI\_TaxID=9606;  
 XN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).

MEDLINE=98054467; PubMed=2824146;  
 Schedlich L.J., Bennetts B.H., Morris B.J.;  
 "Primary structure of a human glandular kallikrein gene.";  
 DNA 6:429-437(1987).  
 [2]  
 SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).  
 TISSUE=Prostate;  
 MEDLINE=92324494; PubMed=1726490;  
 Riegman P.H., Vlietstra R.J., der Korpuit H.A., Romijn J.C.,  
 Trapman J.;  
 "Identification and androgen-regulated expression of two major human  
 glandular kallikrein-1 (hok-1) mRNA species.";  
 Mol. Cell. Endocrinol. 76:181-190(1991).  
 [3]  
 SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.  
 MEDLINE=20012757; PubMed=10544017;  
 Liu X.F., Essand M., Vasmatazis G., Lee B., Pastan I.;  
 "Identification of three new alternate human kallikrein 2  
 evidence of long transcript and alternative splicing.";  
 Biochem. Biophys. Res. Commun. 264:833-839(1999).  
 [4]  
 SEQUENCE FROM N.A.  
 MEDLINE=20510030; PubMed=11054574;  
 Gan L., Lee I., Smith R., Argonza-Barrett R., Lei H., McCuaig J.,  
 Moss P., Paepier B., Wang K.;  
 "Sequencing and expression analysis of the serine protease gene  
 cluster located in chromosome 19q13 region.";  
 Gene 257:119-130(2000).  
 [5]  
 SEQUENCE FROM N.A.  
 MEDLINE=20510030; PubMed=11054574;  
 Auerbach J.E., McCready P.M., Skowronski E., Viswanathan V.,  
 Burkhardt-Schultz K.J., Gordon L., Dias J., Ramirez M., Stilwagen S.,  
 Phan H., Velasco N., Do L., Regala W., Terry A., Ganes J.,  
 Danganan L., Erler A., Christensen M., Georgescu A., Avila J., Liu S.,  
 Duarte S., Andreise T., Trankheim M., Amico-Keller G., Coefield J.,  
 Arellano A., Lucas S., Bruce R., Thomas P., Quan G., Kronmiller B.,  
 Kobayashi A., Olsen A.S., Carraro A.V.;  
 "Sequence analysis of a 4.8 Mb region of 19q13.4 between KLK1 and  
 SYT3.";  
 Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
 [6]  
 SEQUENCE FROM N.A. (ISOFORM 1).  
 TISSUE=Prostate;  
 MEDLINE=22388257; PubMed=12477932;  
 Strausberg R.B., Feingold E.A., Grouse L.H., Derge J.G., Schuler G.D.,  
 Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 Brownstein M.J., Ustin B., Tothiyuki S., Carninci P., Prange C.,  
 Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
 Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
 Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 Butterfield J.S.N., Kravinsky M.I., Skalska U., Smailus D.E.,  
 Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 "Generation and initial analysis of more than 15,000 full-length  
 human and mouse cDNA sequences.";  
 Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 [7]  
 VARIANT LEU-18.  
 MEDLINE=99318093; PubMed=10391209;  
 Cargill M., Altschuler D., Ireland J., Sklar P., Ardlie K., Patil N.,  
 Shaw N., Lane C.R., Lim E.P., Kalyanaram N., Nemesh J., Ziaugra L.,  
 Friedland L., Rolfe A., Warrington J., Lipshutz R., Daley G.Q.,  
 Lander E.S.;  
 "Characterization of single-nucleotide polymorphisms in coding regions

of human Genes.";  
 Nat. Genet. 22:231-238(1999).  
 [8]  
 ERRATUM.  
 RA Cargill M., Altschuler D., Ireland J., Sklar P., Ardlie K., Patil N.,  
 Shaw N., Lane C.R., Lim E.P., Kalyanaram N., Nemesh J., Ziaugra L.,  
 Friedland L., Rolfe A., Warrington J., Lipshutz R., Daley G.Q.,  
 Lander E.S.;  
 Nat. Genet. 23:373-373(1999).  
 CC -1- FUNCTION: Glandular kallikreins cleave Met-Lys and Arg-Ser bonds  
 in kininogen to release Lys-bradykinin.  
 CC -1- CATALYTIC ACTIVITY: Preferential cleavage of Arg-|-Xaa bonds in  
 small molecule substrates. Highly selective action to release  
 kallidin (lysyl-bradykinin) from kininogen involves hydrolysis of  
 Met-|-Xaa or Leu-|-Xaa.  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=3;  
 CC Name=1;  
 CC IsoId=P20151-1; Sequence=Displayed;  
 CC Name=2; Synonyms=PGK-10A;  
 CC IsoId=P20151-2; Sequence=VSP\_005399;  
 CC Name=3;  
 CC IsoId=P20151-3; Sequence=VSP\_005400;  
 CC -1- SIMILARITY: Belongs to peptidase family S1. Kallikrein subfamily.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; M18156; -; NOT ANNOTATED\_CDS.  
 CC EMBL; M18157; AAA74454.1; -;  
 CC EMBL; S39329; AAD13816.1; -;  
 CC EMBL; S39323; AAD13817.1; -;  
 CC EMBL; AF188745; AAF08275.1; -;  
 CC EMBL; AF188746; AAF08276.1; -;  
 CC EMBL; AF189747; AAF08277.1; -;  
 CC EMBL; AF243527; AAG33356.1; -;  
 CC EMBL; AC037199; -; NOT ANNOTATED\_CDS.  
 CC EMBL; BC005196; AAH05196.1; -;  
 CC PIR; A29586; A29586.  
 CC HSP; P07288; 1PFA.  
 CC MEROPS; S01.161; -;  
 CC Genew; HGNC:6383; KLK2.  
 CC MIM; 147960; -;  
 CC GO; GO:0004293; F.tissue kallikrein activity; TAS.  
 CC InterPro; IPR009003; Cys Ser trypsin.  
 CC InterPro; IPR001254; Peptidase S1.  
 CC InterPro; IPR001314; Peptidase\_S1A.  
 CC Pfam; PF00089; trypsin; 1.  
 CC PRINTS; PR00722; CHYMOTRYPSIN.  
 CC SMART; SM00020; Tryp\_SPC; 1.  
 CC PROSITE; PS00240; TRYPsin DOM; 1.  
 CC PROSITE; PS00134; TRYPsin HS; 1.  
 CC PROSITE; PS00135; TRYPsin SER; 1.  
 CC Hydrolase; Serine protease; Glycoprotein; Multigene family; Zymogen;  
 KW Signal; Alternative splicing; Polymorphism.  
 FT SIGNAL 1 18 PROBABLE.  
 FT PROPEP 19 24 ACTIVATION PEPTIDE (PROBABLE).  
 FT CHAIN 25 261 GLANDULAR KALLIKREIN 2.  
 FT ACT\_SITE 65 65 CHANGE RELAY SYSTEM.  
 FT ACT\_SITE 120 120 CHANGE RELAY SYSTEM.  
 FT ACT\_SITE 213 213 CHANGE RELAY SYSTEM.  
 FT DISULFID 31 173 BY SIMILARITY.  
 FT DISULFID 50 66 BY SIMILARITY.  
 FT DISULFID 152 219 BY SIMILARITY.  
 FT DISULFID 184 198 BY SIMILARITY.  
 FT DISULFID 209 234 BY SIMILARITY.  
 FT CARBOHYD 102 102 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT VARSPLIC 211 261 GDSGGFLVGVQIGITSGPEPCALPKPAVYTKVHYRK

WIKDTAAMP -> VSHPSQHQLEK (in isoform 2).

VARSPLIC 165 261  
Missing (in isoform 3).  
/FtId=VSP 005400.  
V -> L (in dBSNP:6072).  
/FtId=VAR 014164.  
SEQUENCE 261 AA; 28671 MW; 9CF7F4A1162EF42D CRC64;

Query Match 80.0%; Score 40; DB 1; Length 261;  
Best Local Similarity 77.8%; Pred. No. 1.6;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

1 GPLVCRGTL 9  
|||||  
215 GPLVCGVL 223

## RESULT 7

AK3 HUMAN STANDARD; PRT; 261 AA.  
C P07288; Q16272;  
I 01-APR-1988 (Rel. 07, Created)  
I 11-JUL-1989 (Rel. 11, Last sequence update)  
I 15-MAR-2004 (Rel. 13, Last annotation update)  
E Prostate specific antigen precursor (EC 3.4.21.77) (PSA) (Gamma-seminoprotein) (Kallikrein 3) (seminogelase) (Seminin) (P-30 antigen).  
N KLK3 OR APS.  
S Homo sapiens (Human).  
C Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
C Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
X NCBI TaxID=9606;  
N P SEQUENCE FROM N.A.  
C TISSUE=Prostate; PubMed=2467258;  
X MEDLINE=92183632; PubMed=2467258;  
A Digby M.R., Zhang X.Y., Richard R.I.;  
T "Human prostate specific antigen (PSA) gene: structure and linkage to the kallikrein-like gene, hGK-1";  
I Nucleic Acids Res. 17:2137-2137(1989).  
N [2]  
P SEQUENCE FROM N.A.  
X MEDLINE=89282407; PubMed=2471958;  
A Klobeck H.-G., Combratio G., Schulz P., Arbusow V., Fittler F.;  
T "Genomic sequence of human prostate specific antigen (PSA)";  
I Nucleic Acids Res. 17:3981-3981(1989).  
N [3]  
P SEQUENCE FROM N.A.  
C TISSUE=Leukocyte;  
X MEDLINE=89302090; PubMed=2472789;  
A Lundwall A.;  
T "Characterization of the gene for prostate-specific antigen, a human glandular kallikrein";  
I Biochem. Biophys. Res. Commun. 161:1151-1159(1989).  
N [4]  
P SEQUENCE FROM N.A.  
C TISSUE=Prostate;  
X MEDLINE=87190978; PubMed=2436946;  
A Lundwall A., Lilja H.;  
T "Molecular cloning of human prostate specific antigen cDNA";  
I FEBS Lett. 214:317-322(1987).  
N [5]  
P SEQUENCE FROM N.A.  
C TISSUE=Prostate;  
X MEDLINE=89246551; PubMed=2470373;  
A Henttu P., Viikio P.;  
T "cDNA coding for the entire human prostate specific antigen shows high homologies to the human tissue kallikrein genes";  
I Biochem. Biophys. Res. Commun. 160:903-910(1989).  
N [6]  
P SEQUENCE FROM N.A.  
C TISSUE=Prostate;  
X MEDLINE=89165891; PubMed=2466464;

RA Riegman P.H.J., Vlietstra R.J., van der Korput J.A.G.M., Romijn J.C.,  
RT Chapman J.;  
RL "Characterization of the prostate-specific antigen gene: a novel human kallikrein-like gene";  
RN Biochem. Biophys. Res. Commun. 159:95-102(1989).  
RP [7]  
P SEQUENCE FROM N.A.  
C TISSUE=Prostate;  
X MEDLINE=20510030; PubMed=11054574;  
A Gan L., Lee I., Smith R., Argonza-Barrett R., Lei H., McCuaig J.,  
T Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.  
N [8]  
P SEQUENCE FROM N.A.  
C MEDLINE=20510030; PubMed=11054574;  
A Moss P., Paepker B., Wang K.;  
T "Sequencing and expression analysis of the serine protease gene cluster located in chromosome 19q13 region";  
N Gene 257:119-130(2000).  
RP [9]  
P SEQUENCE FROM N.A.  
C TISSUE=PNS, and Prostate;  
X MEDLINE=22388257; PubMed=12477932;  
A Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
T Strausberg R.L., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
N Klausner R.D., Buetow K.H., Schaefer C.F., Bhat N.K.,  
P Altschul S.F., Zeeberg B., Moore T., Max S.I., Wang J., Hsieh F.,  
T Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
N Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
P Scapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
T Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci F., Prange C.,  
N Rana S.S., Moquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,  
P Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
T Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
N Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
P Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
T Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
N Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
P Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
T Butterfield Y.C., Krzywinski M.T., Skalska U., Smailus D.E.,  
N Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
P "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences";  
N Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RP [10]  
P SEQUENCE OF 1-176 FROM N.A.  
X MEDLINE=95079406; PubMed=7527295;  
A Monne M., Croce C.M., Yu H., Diamandis E.P.;  
T "Molecular characterization of prostate-specific antigen messenger RNA expressed in breast tumors";  
N Cancer Res. 54:6344-6347(1994).  
RP [11]  
P SEQUENCE OF 17-261 FROM N.A.  
X MEDLINE=88289366; PubMed=2456523;  
A Schulz P., Stucka R., Feldmann H., Combratio G., Klobeck H.-G.,  
T Fittler F.;  
N "Sequence of a cDNA clone encompassing the complete mature human prostate specific antigen (PSA) and an unspliced leader sequence";  
N Nucleic Acids Res. 16:6226-6226(1988).  
RP [12]  
P SEQUENCE OF 25-261.  
X MEDLINE=86205857; PubMed=2422647;  
A Watt K.W.K., Lee P.J., M'Timkulu T., Chan W.P., Loor R.;  
T "Human prostate-specific antigen: structural and functional similarity with serine proteases";  
N Proc. Natl. Acad. Sci. U.S.A. 83:3166-3170(1986).  
RP [13]  
P SEQUENCE OF 25-261.  
X MEDLINE=88082806; PubMed=3691515;  
A Schaller J., Akiyama K., Tsuda R., Hara M., Marti T., Rickli E.E.;  
T "Isolation, characterization and amino-acid sequence of gamma-semioproten, a glycoprotein from human seminal plasma";  
N Eur. J. Biochem. 170:111-120(1987).  
RP [14]  
P 3D-STRUCTURE MODELING.

MEDLINE=95218633; PubMed=7535613;  
 Villoutreix B.O., Getzoff E.D., Griffin J.H.;  
 "A structural model for the prostate disease marker, human prostate-specific antigen.";  
 Protein Sci. 3:2033-2044 (1994).  
 [15]  
 3D-STRUCTURE MODELING.  
 MEDLINE=98427950; PubMed=9751643;  
 Combs G.S., Bergstrom R.C., Pelleguer J.L., Baker S.I., Navre M.,  
 Smith M.M., Tainer J.A., Madison E.L., Corey D.R.;  
 "Substrate specificity of prostate-specific antigen (PSA).";  
 Chem. Biol. 5:475-488(1998).  
 -!- FUNCTION: Presumably hydrolyze the high molecular mass seminal vesicle protein thus leading to the liquefaction of the seminal coagulum.  
 -!- CATALYTIC ACTIVITY: Preferential cleavage: Tyr-|-Xaa.  
 -!- SIMILARITY: Belongs to peptidase family S1. Kallikrein subfamily.  
 -----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 R EMBL; X14810; CRA32915.1; -;  
 R EMBL; X13940; CRA32123.1; -;  
 R EMBL; X13941; CRA32124.1; ALT\_SEQ.  
 R EMBL; X13942; CAB46487.1; -;  
 R EMBL; X13943; CRA32126.1; -;  
 R EMBL; X13944; CRA32127.1; -;  
 R EMBL; X05332; CRA28947.1; -;  
 R EMBL; X07730; -; NOT\_ANNOTATED\_CDS.  
 R EMBL; M27274; AAB60192.1; -;  
 R EMBL; M26663; AAB58802.1; -;  
 R EMBL; U17040; AAB56764.1; -;  
 R EMBL; AF243527; AAG33355.1; -;  
 R EMBL; BC005307; AAH05307.1; -;  
 R EMBL; BC050595; AAH05095.2; -;  
 R EMBL; BC056665; AAH56665.1; -;  
 R EMBL; S75755; AAD14185.1; ALT\_INT.  
 R EMBL; M24543; AAB60193.1; ALT\_SEQ.  
 R PIR; A32297; A32297.  
 R PDB; 1EFA; 26-JAN-95.  
 R PDB; 2PSA; 16-MAR-99.  
 R MEROPS; S01.162; -;  
 R Gene; HGNC:6364; KLK3.  
 R MIM; 176820; -;  
 R GO; GO:0008236; F:serine-type peptidase activity; TAS.  
 R InterPro; IPR009003; Cys Ser trypsin.  
 R InterPro; IPR001254; Peptidase S1.  
 R InterPro; IPR001314; Peptidase\_S1A.  
 R Pfam; PF00089; trypsin; 1.  
 R PRINTS; PR00722; CHYMOTRYPSIN.  
 R SMART; SM00020; TRY SP; 1.  
 R PROSITE; PS00240; TRYPSIN\_DOM; 1.  
 R PROSITE; PS00134; TRYPSIN\_HIS; 1.  
 R PROSITE; PS00135; TRYPSIN\_SER; 1.  
 R Hydrolase; Serine protease; Antigen; zymogen; Signal;  
 R 3D-structure. 1  
 R SIGNAL. 18  
 R PROPEP. 19  
 R CHAIN. 25  
 R ACT\_SITE. 65  
 R ACT\_SITE. 120  
 R ACT\_SITE. 213  
 R CARBOHYD. 69  
 R DISULFID. 31  
 R DISULFID. 50  
 R DISULFID. 152  
 R DISULFID. 184  
 R DISULFID. 209

FT CONFLICT 64 64 A -> T (IN REP. 11).  
 Query Match 80.0%; Score 40; DB 1; Length 261;  
 Best Local Similarity 77.8%; Pred. No. 1.6;  
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QV 1 GPLVCRGTL 9  
 Db 215 GPLVCGVL 223  
 ID KKL1\_RAT STANDARD; PRT; 261 AA.  
 AC P00758;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE Glandular kallikrein, pancreatic 1 precursor (EC 3.4.21.35) (Tissue  
 DE kallikrein) (PS kallikrein) (RCK-1).  
 GN KKL1 OR KKL-1.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 CX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=83117659; PubMed=6961406;  
 RA Swift G.H., Dargatzis J.C., Ashley P.L., Cummings S.W., McDonald R.J.;  
 RT "Rat pancreatic kallikrein mRNA: nucleotide sequence and amino acid  
 RT sequence of the encoded preproenzyme.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 79:7263-7267(1982).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX TISSUE-Submaxillary gland;  
 RX MEDLINE=86051477; PubMed=2998455;  
 RA Ashley P.L., McDonald R.J.;  
 RT "Kallikrein-related mRNAs of the rat submaxillary gland: nucleotide  
 RT sequences of four distinct types including tonin.";  
 RL Biochemistry 24:4512-4520(1985).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX TISSUE=Kidney;  
 RX MEDLINE=89327211; PubMed=2753879;  
 RA Inoue H., Fukui K., Miyake Y.;  
 RT "Identification and structure of the rat true tissue kallikrein gene  
 RT expressed in the kidney.";  
 RL J. Biochem. 105:834-840(1989).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89214217; PubMed=2708383;  
 RA Wines D.R., Brady J.M., Pritchett D.B., Roberts J.L., Macdonald R.J.;  
 RT "Organization and expression of the rat kallikrein gene family.";  
 RL J. Biol. Chem. 264:7653-7662(1989).  
 RN [5]  
 RP SEQUENCE OF 48-261 FROM N.A.  
 RX MEDLINE=86131678; PubMed=3004582;  
 RA Gerald W.L., Chao J., Chao L.;  
 RT "Immunological identification of rat tissue kallikrein cDNA and  
 RT characterization of the kallikrein gene family.";  
 RL Biochim. Biophys. Acta 866:1-14(1996).  
 CC -!- FUNCTION: Glandular kallikreins cleave Met-Lys and Arg-Ser bonds  
 CC in kininogen to release Lys-bradykinin.  
 CC -!- CATALYTIC ACTIVITY: Preferential cleavage of Arg-|-Xaa bonds in  
 CC small molecule substrates. Highly selective action to release  
 CC kallidin (lysyl-bradykinin) from kininogen involves hydrolysis of  
 CC Met-|-Xaa or Leu-|-Xaa.  
 CC -!- SIMILARITY: Belongs to peptidase family S1. Kallikrein subfamily.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC at the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```

EMBL; J00758; -; NOT ANNOTATED CDS.
EMBL; M11563; AAA41464.1; ALT_INIT.
EMBL; M23876; AAA41462.1; -.
EMBL; M23874; AAA41462.1; JOINED.
EMBL; M23875; AAA41462.1; JOINED.
EMBL; D00448; BAA00346.1; ALT_INIT.
EMBL; D00446; BAA00346.1; JOINED.
EMBL; D00447; BAA00346.1; JOINED.
EMBL; X03560; CAA27247.1; -.
PIR; A00944; KQTP.
HSP; P00757; ISGF.
MEROPS; S01.405; -.
InterPro; IPR009003; Cys Ser trypsin.
InterPro; IPR001254; Peptidase_S1.
InterPro; IPR001314; Peptidase_S1A.
Pfam; PF00089; trypsin; 1.
PRINTS; PR00722; CHYMOTRYPSIN.
SMART; SM00020; Tryp_Spc; 1.
PROSITE; PS0240; TRYPSIN_DOM; 1.
PROSITE; PS00134; TRYPSIN_HIS; 1.
PROSITE; PS00135; TRYPSIN_SER; 1.
Hydrolase; Serine protease; Zymogen;
Signal.
T SIGNAL 1 18 PROBABLE.
T PROPEP 19 24 ACTIVATION PEPTIDE (PROBABLE).
T CHAIN 25 261 GLANDULAR KALLIKREIN 1.
T CHAIN 25 111 CHAIN 1.
T CHAIN 112 261 CHAIN 2.
T ACT_SITE 65 65 CHARGE RELAY SYSTEM.
T ACT_SITE 120 120 CHARGE RELAY SYSTEM.
T ACT_SITE 213 213 CHARGE RELAY SYSTEM.
T DISULFID 31 173 BY SIMILARITY.
T DISULFID 50 66 BY SIMILARITY.
T DISULFID 152 219 BY SIMILARITY.
T DISULFID 184 198 BY SIMILARITY.
T DISULFID 209 234 BY SIMILARITY.
T CARBOHYD 108 108 N-LINKED (GLCNAC... ) (PROBABLE).
Q SEQUENCE 261 AA; F2F99C0227A7882B CRC64;

Query Match 78.0%; Score 39; DB 1; Length 261;
Best Local Similarity 66.7%; Pred. No. 2.5;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Y 1 GPLVCRGTL 9
|||:|
b 215 GPLICNGVL 223

RESULT 9
LKLC HUMAN STANDARD; PRT; 248 AA.
C QUKRO; QUKR1;
Y 16-OCT-2001 (Rel. 40, Created)
Y 16-OCT-2001 (Rel. 40, Last sequence update)
Y 16-OCT-2001 (Rel. 40, Last annotation update)
Y Kallikrein 12 precursor (EC 3.4.21.-) (Kallikrein-like protein 5)
(EK-L5).
Y KLIK2 OR KLIK5.
Y Homo sapiens (Human).
Y Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Y Mammalia; Eutheria; Primates; Craniata; Hominidae; Homo.
Y NCBI_TaxID=9606;
Y [1]
Y SEQUENCE FROM N.A. (ISOFORM 1).
Y MEDLINE=20118156; PubMed=10652563;
Y Yousef G.M., Luo L.-Y., Diamandis E.P.;
Y "Identification of novel human kallikrein-like genes on chromosome
Y 19q13.3-q13.4."
Y Anticancer Res. 19:2843-2852 (1999).

```

```

[2] SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RP Yousef G.M., Magklara A., Scorillas A., Diamandis E.P.;
RA "Cloning of new alternatively spliced forms of the kallikrein-like
RT gene 5 (KLK-L5).";
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
[3]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE=20510030; PubMed=11054574;
RA Gan L., Lee I., Smith R., Argonza-Barrett R., Lei H., McCuaig J.,
RA Moss P., Paepel B., Wang K.;
RA "Sequencing and expression analysis of the serine protease gene
RT cluster located in chromosome 19q13 region.";
RL Gene 257:119-130 (2000).
[4]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RA Lamerdin J.E., McCreedy P.M., Skowronski E., Viswanathan V.,
RA Burkhardt-Schultz K., Gordon L., Dias J., Ramirez M., Stilwgen S.,
RA Phan H., Velasco N., Do L., Regala W., Terry A., Brower A., Ganes J.,
RA Danganan L., Erlar A., Christensen M., Georgescu A., Avila J., Liu S.,
RA Andreise T., Frankheim M., Attix C., Amico-Keller G., Coefield J.,
RA Duarte S., Lucas S., Bruce R., Thomas P., Quan G., Kronmiller B.,
RA Arellano A., Sanders C., Ow D., Nolan M., Trong S., Kobayashi A.,
RA Olsen A.S., Carrano A.V.;
RL "Sequence analysis of chromosome 19q13.4.";
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Secreted (Probable).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=Q9UKR0-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q9UKR0-2; Sequence=VSP_005403;
CC -!- SIMILARITY: Belongs to peptidase family S1. Kallikrein subfamily.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation at
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announcement/
or send an email to license@isb-sib.ch)
-----
EMBL; AF135025; AAD26426.2; -.
EMBL; AF135025; AAF06065.1; -.
EMBL; AF243527; AAG33365.1; -.
EMBL; AC011473; AAG32358.1; -.
HSP; P00763; LDPO.
MEROPS; S01.020; -.
Genew; HGNC:6360; KLK12.
MIM; 605539; -.
GO; GO:0005576; C:extracellular; NAS.
GO; GO:0004252; F:serine-type endopeptidase activity; NAS.
GO; GO:0006508; P:proteolysis and peptidolysis; NAS.
InterPro; IPR009003; Cys Ser trypsin.
InterPro; IPR001254; Peptidase_S1.
InterPro; IPR001314; Peptidase_S1A.
Pfam; PF00089; trypsin; 1.
PRINTS; PR00722; CHYMOTRYPSIN.
SMART; SM00020; Tryp_Spc; 1.
PROSITE; PS0240; TRYPSIN_DOM; 1.
PROSITE; PS00134; TRYPSIN_HIS; 1.
PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Glycoprotein; Signal;
KW Alternative splicing.
FT SIGNAL 1 17 POTENTIAL.
FT CHAIN 18 248 KALLIKREIN 12.
FT ACT_SITE 62 62 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 108 108 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 200 200 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 28 161 BY SIMILARITY.
FT DISULFID 47 63 BY SIMILARITY.
FT DISULFID 133 235 BY SIMILARITY.

```

T DISULFID 140 206 BY SIMILARITY.  
T DISULFID 172 186 BY SIMILARITY.  
T DISULFID 196 222 BY SIMILARITY.  
T CARBOHYD 24 24 N-LINKED (GLCNAC. . .) (POTENTIAL).  
T CARBOHYD 163 163 N-LINKED (GLCNAC. . .) (POTENTIAL).  
T VARSPLIC 236 248 KYDVMRTMTRNN -> NSTVLGLGTSWNFNSCQPF (in isoform 2).  
T /PTID=VSP\_005403.  
T  
Q SEQUENCE 248 AA; 26733 MW; BB473E95F8BAF703 CRC64;  
Query Match 76.0%; Score 38; DB 1; Length 248;  
Best Local Similarity 77.8%; Pred. No. 3.7;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Y 1 GPLVCRGTL 9  
| | | | |  
b 202 GPLVCGGV 210  
RESULT 10  
D\_FAD HUMAN STANDARD; PRT; 253 AA.  
C P00746; 1986 (Rel. 01, Created)  
T 21-JUL-1986 (Rel. 37, Last sequence update)  
T 15-DEC-1998 (Rel. 43, Last annotation update)  
T 15-MAR-2004 (Rel. 43, Last annotation update)  
E Complement factor D precursor (EC 3.4.21.46) (C3 convertase activator)  
E (Properdin factor D) (Adipsin).  
N DF.  
N Homo sapiens (Human).  
S Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
M Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
X NCBI\_TaxID=9606;  
X [1]  
P SEQUENCE OF 8-253 FROM N.A.  
X MEDLINE=92250520; PubMed=1374388;  
X White R.T., Damm D., Hancock N., Rosen B.S., Lowell B.B., Usher P.,  
X Flier J.S., Spiegelman B.M.;  
X "Human adipon is identical to complement factor D and is expressed  
X at high levels in adipose tissue.";  
X J. Biol. Chem. 267:9210-9213(1992).  
X [2]  
P SEQUENCE FROM N.A.  
X Flier J.S., Spiegelman B.M., Rosen B.M.;  
X Patent number WO9006365, 14-JUN-1990.  
X [3]  
P SEQUENCE OF 26-252.  
X MEDLINE=85000441; PubMed=6383466;  
X Niewann M.A., Bhowan A.S., Bennett J.C., Volanakis J.E.;  
X "Amino acid sequence of human D of the alternative complement  
X pathway.";  
X Biochemistry 23:2482-2486(1984).  
X [4]  
P PARTIAL SEQUENCE OF 26-252.  
X MEDLINE=84108950; PubMed=6363133;  
X Johnson D.M.A., Gagnon J., Reid K.B.M.;  
X "Amino acid sequence of human factor D of the complement system.  
X Similarity in sequence between factor D and proteases of non-plasma  
X origin.";  
X FEBS Lett. 166:347-351(1984).  
X [5]  
P PARTIAL SEQUENCE OF 26-61 AND 194-220.  
X MEDLINE=84256515; PubMed=6821372;  
X Johnson D.M.A., Gagnon J., Reid K.B.M.;  
X "Factor D of the alternative pathway of human complement.  
X Purification, alignment and N-terminal amino acid sequences of the  
X major cyanogen bromide fragments, and localization of the serine  
X residue at the active site.";  
X Biochem. J. 187:863-874(1980).  
X [6]  
P PARTIAL SEQUENCE OF 26-82.  
X MEDLINE=80145719; PubMed=6987665;  
X Volanakis J.E., Bhowan A.S., Bennett J.C., Mole J.E.;  
X "Partial amino acid sequence of human factor D:homology with serine  
X proteases.";  
X Proc. Natl. Acad. Sci. U.S.A. 77:1116-1119(1980).  
X [7]  
P PARTIAL SEQUENCE OF 26-78.  
X MEDLINE=81054886; PubMed=6776531;  
X Davis A.E. III;  
X "Active site amino acid sequence of human factor D.";  
X Proc. Natl. Acad. Sci. U.S.A. 77:4938-4942(1980).  
X [8]  
P X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).  
X MEDLINE=94118317; PubMed=828289;  
X Narayana S.V.L., Carson M., El-Kabbani O., Kilpatrick J.M., Moore D.,  
X Chen X., Bugg C.E., Volanakis J.E., Delucas J.J.;  
X "Structure of human factor D. A complement system protein at 2.0-A  
X resolution.";  
X J. Mol. Biol. 235:695-708(1994).  
X [9]  
P X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).  
X MEDLINE=96025834; PubMed=7592653;  
X Kim S., Narayana S.V., Volanakis J.E.;  
X "Crystal structure of a complement factor D mutant expressing  
X enhanced catalytic activity.";  
X J. Biol. Chem. 270:24399-24405(1995).  
X -!- FUNCTION: Factor D cleaves factor B when the latter is complexed  
X with factor C3b, activating the C3bbb complex, which then becomes  
X the C3 convertase of the alternate pathway. Its function is  
X homologous to that of C1s in the classical pathway.  
X -!- CATALYTIC ACTIVITY: Cleaves component factor B (Arg-|-Lys) when in  
X complex with C3b or with cobra venom factor (CVF).  
X -!- SIMILARITY: Belongs to peptidase family S1.  
X -!- CAUTION: In addition to the conflicts shown in the feature table,  
X Ref.3 sequence had a peptide in the wrong order and another one  
X missing.  
X -----  
X This SWISS-PROT entry is copyright. It is produced through a collaboration  
X between the Swiss Institute of Bioinformatics and the EMBL outstation -  
X the European Bioinformatics Institute. There are no restrictions on its  
X use by non-profit institutions as long as its content is in no way  
X modified and this statement is not removed. Usage by and for commercial  
X entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
X or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
X -----  
X EMBL; M84526; AAA35527.1; ALT\_INIT.  
X F1; A40197; DBHU.  
X PDB; 1DFP; 25-FEB-98.  
X PDB; 1DST; 11-JUL-96.  
X PDB; 1DSU; 17-AUG-96.  
X PDB; 1BIO; 22-JUN-99.  
X PDB; 1DIC; 22-JUL-99.  
X PDB; 1FDP; 03-DEC-99.  
X PDB; 1HFD; 22-JUN-99.  
X MEROPS; S01.191; -.  
X Genew; HGNC:2771; DF.  
X MIM; 143350; -.  
X GO; GO:0003817; F:complement factor D activity; TAS.  
X GO; GO:0008236; F:serine-type peptidase activity; TAS.  
X GO; GO:0006956; P:complement activation; TAS.  
X GO; GO:0006508; P:proteolysis and peptidolysis; TAS.  
X InterPro; IPR009003; Cys Ser trypsin.  
X InterPro; IPR001254; Peptidase S1.  
X InterPro; IPR001314; Peptidase\_S1A.  
X Pfam; PF00089; trypsin; 1.  
X PRINTS; PR00722; CHYMOTRYPSIN.  
X SMART; SM00020; TRY SPG; 1.  
X PROSITE; PS00240; TRYPSIN\_DOM; 1.  
X PROSITE; PS00134; TRYPSIN\_HIS; 1.  
X PROSITE; PS00135; TRYPSIN\_SER; 1.  
X Complement alternate pathway; Plasma; Hydrolase; Serine protease;  
X Zymogen; Glycoprotein; Signal; 3D-structure.  
X SIGNAL 1 20 POTENTIAL.  
X PROPEP 21 25 ACTIVATION PEPTIDE (POTENTIAL).  
X CHAIN 26 253 COMPLEMENT FACTOR D.  
X FT



```
ACT SITE 66 CHARGE RELAY SYSTEM.
ACT SITE 112 CHARGE RELAY SYSTEM.
ACT SITE 208 CHARGE RELAY SYSTEM.
DISULFID 51
DISULFID 148
DISULFID 179
DISULFID 204
CONFLICT 26
CONFLICT 35
CONFLICT 40
CONFLICT 49
CONFLICT 52
CONFLICT 59
CONFLICT 63
CONFLICT 73
CONFLICT 83
CONFLICT 84
CONFLICT 94
CONFLICT 96
CONFLICT 136
CONFLICT 178
CONFLICT 243
CONFLICT 250
CONFLICT 250
CONFLICT 250
STRAND 27
TURN 28
STRAND 30
TURN 31
TURN 34
TURN 38
STRAND 40
TURN 46
STRAND 48
TURN 58
STRAND 60
TURN 65
TURN 66
TURN 76
STRAND 84
TURN 90
TURN 93
TURN 104
TURN 112
STRAND 116
STRAND 127
TURN 128
STRAND 130
TURN 144
TURN 147
TURN 158
TURN 165
TURN 167
TURN 176
TURN 183
TURN 191
STRAND 193
STRAND 202
STRAND 204
TURN 205
TURN 208
TURN 211
TURN 215
STRAND 217
STRAND 226
TURN 229
STRAND 236
TURN 242
TURN 245
SEQUENCE 253 AA; BD553B70BD55C6AD CRC64;

Query Match 76.0%; Score 38; DB 1; Length 253;
Best Local Similarity 77.8%; Pred. No. 3.8;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

I -> M (IN REF. 1).
H -> F (IN REF. 6).
M -> V (IN REF. 6).
H -> E (IN REF. 4 AND 5).
G -> A (IN REF. 1 AND 2).
Q -> R (IN REF. 1 AND 2).
S -> T (IN REF. 4).
D -> G (IN REF. 4).
HSLS -> THLP (IN REF. 3).
HS -> ST (IN REF. 4).
MISSING (IN REF. 4).
D -> E (IN REF. 4).
Q -> G (IN REF. 4).
TCNRTHDGAITE -> KCRLYDVL (IN REF. 4).
S -> T (IN REF. 3).
S -> H (IN REF. 3).
MISSING (IN REF. 4).

PRN3 MOUSE
ID PRN3 MOUSE STANDARD; PRT; 254 AA.
AC Q61096; Q08809;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Myeloblastin precursor (EC 3.4.21.76) (Proteinase 3) (PR-3).
GN PRN3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SvJ;
RX MEDLINE=99126347; PubMed=9925946;
RA Sturrock A., Franklin K.F., Wu S.-Q., Hoidal J.R.;
RT "Characterization and localization of the genes for mouse proteinase-3
RT (Prtn3) and neutrophil elastase (Ela2).";
PL Cytogenet. Cell Genet. 83:104-108(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SvJ;
RX MEDLINE=97330929; PubMed=9187364;
RA Jenne D.E., Froehlich L., Hummel A.M., Specks U.;
RT "Cloning and functional expression of the murine homologue of
RT proteinase 3: implications for the design of murine models of
RT vasculitis.";
RL FEBS Lett. 408:187-190(1997).
RN [3]
RP SEQUENCE OF 2-254 FROM N.A.
RC STRAIN=BALE/C;
RX MEDLINE=97362044; PubMed=9211743;
RA Aveskogh M., Lutzelshwab C., Huang M.R., Hellman L.;
RT "Characterization of cDNA clones encoding mouse proteinase 3
RT (myeloblastin) and cathepsin G.";
RL Immunogenetics 46:181-191(1997).
CC -!- FUNCTION: Polymorphonuclear leukocyte serine protease that
CC degrades elastin, fibronectin, laminin, vitronectin, and collagen
CC types I, III, and IV (in vitro) and causes emphysema when
CC administered by tracheal insufflation to hamsters (By similarity).
CC -!- CATALYTIC ACTIVITY: Hydrolysis of proteins, including elastin, by
CC preferential cleavage: Ala-Xaa > Val-Xaa.
CC -!- SIMILARITY: Belongs to peptidase family S1. Elastase subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF082186; AAC79701.1; -
CC EMBL; U97073; AAB58055.1; -
CC EMBL; U43525; AAB6721.1; -
CC HSP; P24158; 1FUU.
CC MEROPS; S01.134; -.
CC MGD; MGI:893580; Prtn3.
CC InterPro; IPR009003; Cys Ser trypsin.
CC InterPro; IPR001254; Peptidase S1.
CC InterPro; IPR001314; Peptidase_S1A.
CC Pfam; PF00089; trypsin; 1.
CC PRINTS; PR00722; CHYMOTRYPSIN.
CC SMART; SM00020; Tryp_SPC; 1.
CC PROSITE; PS50240; TRYPSIN_DOM; 1.
```

1 PROSITE; PS00134; TRYPSIN HIS; 1.  
 2 PROSITE; PS00135; TRYPSIN SER; 1.  
 3 Collagen degradation; Hydrolase; Serine protease; signal; Zymogen;  
 4 Glycoprotein.  
 5 SIGNAL 1 27 BY SIMILARITY.  
 6 CHAIN 28 29 MYELOBLASTIN.  
 7 PROPEP 30 250  
 8 ACT SITE 251 254 BY SIMILARITY.  
 9 ACT SITE 73 73 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 10 ACT SITE 120 120 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 11 ACT SITE 205 205 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 12 CARBOHYD 127 127 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 13 CARBOHYD 176 176 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 14 DISULFID 58 74 BY SIMILARITY.  
 15 DISULFID 154 211 BY SIMILARITY.  
 16 DISULFID 184 190 BY SIMILARITY.  
 17 DISULFID 201 226 BY SIMILARITY.  
 18 CONFLICT 2 2 S -> A (IN REF. 2).  
 19 SEQUENCE 254 AA; 27626 MW; 00CE8989A3CB79CA CRC64;  
 20  
 Query Match 76.0%; Score 38; DB 1; Length 254;  
 Best Local Similarity 66.7%; Pred. No. 3.8;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 Y 1 GPLVCRGTL 9  
 ||||:||||  
 b 207 GPLICNGIL 215  
 ||||:||||  
 RESULT 12  
 SP3 TRIGA STANDARD; PRT; 258 AA.  
 C O13063;  
 T 16-OCT-2001 (Rel. 40, Created)  
 T 16-OCT-2001 (Rel. 40, Last sequence update)  
 T 28-FEB-2003 (Rel. 41, Last annotation update)  
 E Venom serine proteinase 3 precursor (EC 3.4.21.-).  
 N TLG3.  
 S Trimeresurus gramineus (Indian green tree viper) (Green habu snake).  
 C Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 C Lepidodactylia; Squamata; Scleroglossa; Serpentes; Colubroidea;  
 C Viperae; Crotalinae; Trimeresurus.  
 X NCBI\_TaxID=8767;  
 [1]  
 P SEQUENCE FROM N.A.  
 R TISSUE=Venom gland;  
 X MEDLINE=97096898; PubMed=8941719;  
 A Deshigawara M., Ozawa T., Nakashima K., Nobuhisa I., Chijiwa T.,  
 A Shimochigashi Y., Fukumaki Y., Niwa M., Yamashina I., Hattori S.,  
 A Ohno M.;  
 T "Accelerated evolution of crotalinae snake venom gland serine  
 proteases.";  
 X FEBS Lett. 397:83-88(1996).  
 C -!- FUNCTION: Thrombin-like snake venom serine protease.  
 C -!- SUBCELLULAR LOCATION: Secreted.  
 X -!- TISSUE SPECIFICITY: Expressed by the venom gland.  
 C -!- SIMILARITY: Belongs to peptidase family S1. Snake venom subfamily.  
 C  
 C This SWISS-PROT entry is copyright. It is produced through a collaboration  
 C between the Swiss Institute of Bioinformatics and the EMBL outstation  
 C the European Bioinformatics Institute. There are no restrictions on its  
 C use by non-profit institutions as long as its content is in no way  
 C modified and this statement is not removed. Usage by and for commercial  
 C entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 C or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 C  
 C EMBL; D67085; BAA19983.1; -  
 C HSSP; P00763; LDPO.  
 C  
 C MEROPS; S01.185; -  
 C InterPro; IPR009003; Cys Ser trypsin.  
 C InterPro; IPR001254; Peptidase S1.  
 C InterPro; IPR001314; Peptidase\_S1A.  
 C Pfam; PF00089; trypsin; 1.  
 C

DR PRINTS; PRO0722; CHYMOTRYPSIN.  
 DR SMART; SM00020; TRYPC\_SPC; 1.  
 DR PROSITE; PS00240; TRYPSIN DOM; 1.  
 DR PROSITE; PS00134; TRYPSIN HIS; 1.  
 DR PROSITE; PS00135; TRYPSIN SER; 1.  
 KW Hydrolase; Serine protease; Glycoprotein; Zymogen; Signal.  
 FT SIGNAL 1 18 BY SIMILARITY.  
 FT PROPEP 19 24 BY SIMILARITY.  
 FT CHAIN 25 258 VENOM SERINE PROTEINASE 3.  
 FT ACT SITE 65 65 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT SITE 110 110 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT SITE 204 204 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT DISULFID 31 163 BY SIMILARITY.  
 FT DISULFID 50 66 BY SIMILARITY.  
 FT DISULFID 98 256 BY SIMILARITY.  
 FT DISULFID 142 210 BY SIMILARITY.  
 FT DISULFID 174 189 BY SIMILARITY.  
 FT DISULFID 200 225 BY SIMILARITY.  
 FT CARBOHYD 44 44 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 103 103 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 117 117 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 121 121 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 154 154 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 251 251 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 258 AA; 28034 MW; 3160379F61E9F74B CRC64;  
 20  
 Query Match 76.0%; Score 38; DB 1; Length 258;  
 Best Local Similarity 66.7%; Pred. No. 3.8;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 Qy 1 GPLVCRGTL 9  
 ||||:||||  
 Db 206 GPLICNGQL 214  
 ||||:||||  
 RESULT 13  
 K1K8 HUMAN STANDARD; PRT; 260 AA.  
 ID O60259; O9HCB3; Q9U1L9; Q9U047;  
 AC 15-JUL-1999 (Rel. 38, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Neuropilin precursor (EC 3.4.21.-) (NP) (Kallikrein 8) (Ovasin) (Serine  
 DE protease TADG-14) (Tumor-associated differentially expressed gene-14  
 DE protein).  
 DE K1K8 OR PRSS19 OR TADG14 OR NRPN.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OK NCBI\_TaxID=9606;  
 [1]  
 RN SEQUENCE FROM N.A. (ISOFORM 1).  
 RC TISSUE=Hippocampus;  
 RX MEDLINE=98372070; PubMed=9714609;  
 RA Yoshida S., Taniguchi M., Hirata A., Shiosaka S.;  
 RT "Sequence analysis and expression of human neuropilin cDNA and gene.";  
 RL Gene 213:9-16(1998).  
 [2]  
 RN SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).  
 RC TISSUE=Brain;  
 RX MEDLINE=99203457; PubMed=10102990;  
 RA Mitsui S., Tsuruoka N., Yamashiro K., Nakazato H., Yamaguchi N.;  
 RT "A novel form of human neuropilin, a brain-related serine protease, is  
 RT generated by alternative splicing and is expressed preferentially in  
 RT human adult brain.";  
 RL Eur. J. Biochem. 260:627-634(1999).  
 [3]  
 RN SEQUENCE FROM N.A. (ISOFORM 1).  
 RC TISSUE=Ovary;  
 RX MEDLINE=99413504; PubMed=10485494;  
 RA Underwood L.J., Tanimoto H., Wang Y., Shigenasa K., Parmley T.H.,  
 RA O'Brien T.J.;  
 RT "Cloning of tumor-associated differentially expressed gene-14, a novel

serine protease overexpressed by ovarian carcinoma.",  
Cancer Res. 59:4435-4439(1999).  
[4]  
SEQUENCE FROM N.A. (ISOFORM 1).  
Gan L., Gellinas R., Gown A.M., Moss P., Smith R., Wang K.,  
"Molecular cloning and characterization of a novel serine protease,  
ovasin, a potential molecular marker for ovarian carcinomas.",  
Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.  
[5]  
SEQUENCE FROM N.A. (ISOFORM 1).  
MEDLINE=20510030; PubMed=11054574;  
Gan L., Lee I., Smith R., Argonza-Barrett R., Lei H., McQuaig J.,  
Moss P., Paepel B., Wang K.,  
"Sequencing and expression analysis of the serine protease gene  
cluster located in chromosome 19q13 region.",  
Gene 257:119-130(2000).  
[6]  
SEQUENCE OF 1-164 FROM N.A. (ISOFORM 1).  
Lamerdin J.E., McCready P.M., Skowronski B., Viswanathan V.,  
Burkhardt-Schultz K., Gordon L., Dias J., Ramirez M., Stilwagen S.,  
Phan H., Velasco N., Do L., Regala M., Terry A., Brower A., Carnes J.,  
Danganan L., Eriar A., Christensen M., Georgescu A., Avila J., Liu S.,  
Andreise T., Frankel M., Attix C., Amico-Keller G., Coefield J.,  
Duarte S., Lucas S., Bruce R., Thomas P., Quan G., Kronmiller B.,  
Arellano A., Sanders C., Ow D., Nolan M., Trong S., Kobayashi A.,  
Olson A.S., Carrano A.V.,  
"Sequence analysis of chromosome 19q13.4.",  
Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
[7]  
-!- FUNCTION: Suggested to be involved in kindling epileptogenesis and  
hippocampal plasticity.  
-!- CATALYTIC ACTIVITY: Preferential cleavage: Arg-, Lys-.  
-!- SUBCELLULAR LOCATION: Secreted.  
-!- ALTERNATIVE PRODUCTS:  
Event=Alternative splicing; Named isoforms=2;  
Name=1;  
IsoId=060259-1; Sequence=Displayed;  
Name=2;  
IsoId=060259-2; Sequence=VSP\_005401;  
-!- TISSUE SPECIFICITY: Isoform 1 is predominantly expressed in the  
pancreas while isoform 2 is expressed in adult brain and  
hippocampus. Both forms are also found in fetal brain and  
placenta. Not detected in kidney, spleen, liver and lung.  
-!- SIMILARITY: Belongs to peptidase family S1. Kallikrein subfamily.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; AB009849; BAA28673.1; -  
EMBL; AB012761; BAA28676.1; -  
EMBL; AB010780; BAA88884.1; -  
EMBL; AB008390; BAA82865.1; -  
EMBL; AB008927; BAA82866.1; -  
EMBL; AF055982; AAD56050.1; -  
EMBL; AF095742; AAD29579.1; -  
EMBL; AF095743; AAD29574.1; -  
EMBL; AF243527; AAG33361.1; -  
EMBL; AC011473; AAG23254.1; -  
HSP; O61955; INPM.  
Genew; S01.244; -.  
MIM; 605644; -.  
GO; GO:0008236; F:serine-type peptidase activity; TAS.  
GO; GO:0007399; P:neurogenesis; TAS.  
InterPro; IPR009003; Cys\_Ser\_trypsin.  
InterPro; IPR001254; Peptidase\_S1.  
InterPro; IPR001314; Peptidase\_S1A.  
Pfam; PF00089; trypsin; 1.  
PRINTS; PR00722; CHYMOTRYPSIN.

SMART; SM00020; Tryp\_Spc; 1.  
DR PROSITE; PS00240; TRYPSIN\_DOM; 1.  
DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
KW Hydrolase; Serine protease; Glycoprotein; Zymogen; Signal;  
FT SIGNAL 1 28  
FT PROPEP 29 32  
FT CHAIN 33 260  
FT ACT\_SITE 73 73  
FT ACT\_SITE 120 120  
FT ACT\_SITE 212 212  
FT DISULFID 39 173  
FT DISULFID 58 74  
FT DISULFID 145 246  
FT DISULFID 152 218  
FT DISULFID 184 198  
FT DISULFID 208 233  
FT CARBOHYD 110 110  
FT VARSPLIC 23 23  
FT WRSNPLP2AA (in isoform 2).  
FT FTID=VSP\_005401.  
SQ SEQUENCE 260 AA; 28048 MW; EF4395B58C9E660 CRC64;  
Query Match 76.0%; Score 38; DB 1; Length 260;  
Best Local Similarity 77.8%; Pred. No. 3.9;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 GPLVCRGTL 9  
Db 214 GPLVCDGAL 222  
RESULT 14  
KLK3 MACMU STANDARD; PRT; 261 AA.  
ID KLK3 MACMU  
AC P33619;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Prostate specific antigen precursor (EC 3.4.21.35) (PSA) (Gamma-  
DE semineproteolysin) (Kallikrein 3).  
GN KLK3 OR APS.  
OS Macaca mulatta (Rhesus macaque).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
OC Cercopithecoidea; Macaca.  
OX NCBI\_TaxID=9544;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93363642; PubMed=7689340;  
RA Gauthier E.R., Chapdelaine P., Tremblay R.R., Dube J.Y.;  
RT "Characterization of rhesus monkey prostate specific antigen cDNA.";  
RL Biochim Biophys Acta 1174:207-210(1993).  
CC -!- FUNCTION: PSA presumably hydrolyzes the high molecular mass  
CC seminal vesicle protein thus leading to the liquefaction of  
CC the seminal coagulum. PSA is a glandular kallikrein.  
CC -!- CATALYTIC ACTIVITY: Preferential cleavage: Tyr--Xaa.  
CC -!- SIMILARITY: Belongs to peptidase family S1. Kallikrein subfamily.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X73560; CAA51957.1; -  
DR PIR; S35711; S35711.  
DR HSP; P07288; 1PPA.  
DR MEROPS; S01.162; -.  
DR InterPro; IPR009003; Cys\_Ser\_trypsin.

```

1 InterPro: IPR001254; Peptidase_S1.
2 InterPro: IPR001314; Peptidase_S1A.
3 Pfam: PF00089; trypsin_1.
4 PRINTS: PR00722; CHYMOTRYPSIN.
5 SMART: SMO0020; Tryp_Spc; 1.
6 PROSITE: PS02040; TRYPSIN_DOM; 1.
7 PROSITE: PS00134; TRYPSIN_HIS; 1.
8 PROSITE: PS00135; TRYPSIN_SER; 1.
9 Hydrolyase; Serine protease; Glycoprotein; Antigen; Zymogen; Signal.
10 SIGNAL 1 18
11 PROPEP 19 24
12 CHAIN 25 261
13 ACT_SITE 65 65
14 ACT_SITE 120 120
15 ACT_SITE 213 213
16 DISULFID 31 173
17 DISULFID 50 66
18 DISULFID 152 219
19 DISULFID 184 198
20 DISULFID 209 234
21 CARBOHYD 102 102
22 SEQUENCE 261 AA; 28816 MW; 8525B14515967E5C CRC64;

Query Match 76.0%; Score 38; DB 1; Length 261;
Best Local Similarity 77.8%; Pred. No. 3.9;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

y 1 GPLVCRGTL 9
  |||||
b 215 GPLVCDGVL 223

RESULT 15
LK5 HUMAN
D KLK5 HUMAN STANDARD; PRT; 293 AA.
C Q9V337; OSRBG8;
T 16-OCT-2001 (Rel. 40, Created)
T 16-OCT-2001 (Rel. 40, Last sequence update)
T 15-MAR-2004 (Rel. 43, Last annotation update)
E Kallikrein 5 precursor (EC 3.4.21.-) (Stratum corneum tryptic enzyme)
E (Kallikrein-like protein 2) (KLK-L2).
N KLK5 OR SCTE.
X Homo sapiens (Human).
X Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
X Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
X NCBI_taxid=9606;
[1]
[2] SEQUENCE FROM N.A.
[2] TISSUE=Stratum corneum;
[2] MEDLINE=99445563; PubMed=10514489;
[2] Brattsand M., Egelrud T.;
[2] "Purification, molecular cloning, and expression of a human stratum
[2] corneum trypsin-like serine protease with possible function in
[2] desquamation.";
[2] J. Biol. Chem. 274:30033-30040 (1999).
[2]
[2] SEQUENCE FROM N.A.
[2] MEDLINE=20118156; PubMed=10652563;
[2] Yousef G.M., Luo L.-Y., Diamandis E.P.;
[2] "Identification of novel human kallikrein-like genes on chromosome
[2] 19q13.3-q13.4.";
[2] Anticancer Res. 19:2843-2852 (1999).
[3]
[3] SEQUENCE FROM N.A.
[3] MEDLINE=20510030; PubMed=11054574;
[3] Gan L., Lee I., Smith R., Argonza-Barrett R., Lei H., McCuaig J.,
[3] Moss P., Paepel B., Wang K.;
[3] "Sequencing and expression analysis of the serine protease gene
[3] cluster located in chromosome 19q13 region.";
[3] Gene 257:119-130 (2000).
[4]
[4] SEQUENCE FROM N.A.
[4] TISSUE=Ovary;

```

MEDLINE=22388257; PubMed=12477932;

Strauberg R.L., Colling F.S., Wagner L., Shemen C.M., Schuler G.D., Klausner R.D., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Altschul S.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., Hopkins R.F., Marusina K., Farmer A.A., Rubin G.M., Hong L., Diatchenko L., Scarsone M.B., Bonaldo M.F., Casavant T.L., Scheetz T.B., Stapleton M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C., Brownstein M.J., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J., Raha S.S., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Bosak S.A., Worsley K.C., Hale S., Garcia A.M., Gay L.J., Rulyk S.W., Richards S., Worsley K.C., Sodergren E.J., Lu X., Gibbs R.A., Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalus D.E., Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;

"Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences";

Proc Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).

-!- FUNCTION: May be involved in desquamation.

-!- SUBCELLULAR LOCATION: Secreted.

-!- TISSUE SPECIFICITY: Expressed in skin, breast, brain and testis.

-!- SIMILARITY: Belongs to peptidase family S1. Kallikrein subfamily.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL: AF168768; AAF03101.1; --

EMBL: AF135028; AAD26429.1; --

EMBL: AF243527; AAG33358.1; --

EMBL: BC008036; AAK08036.1; --

HSP: P0763; IDPO.

MEROPS: S01.017; --

Genew; HGNC:6366; KLK5.

MIM: 605643; --

GO: GO:0005615; C:extracellular space; TAS.

GO: GO:0004252; F:serine-type endopeptidase activity; NAS.

GO: GO:0008544; P:epidermal differentiation; TAS.

GO: GO:0006508; P:proteolysis and peptidolysis; NAS.

InterPro: IPR009003; Cys\_ser\_trypsin.

InterPro: IPR001254; Peptidase\_S1.

InterPro: IPR001314; Peptidase\_S1A.

Pfam: PF00089; trypsin; 1.

PRINTS: PR00722; CHYMOTRYPSIN.

SMART: SMO0020; Tryp\_Spc; 1.

PROSITE: PS02040; TRYPSIN\_DOM; 1.

PROSITE: PS00134; TRYPSIN\_HIS; 1.

PROSITE: PS00135; TRYPSIN\_SER; 1.

Hydrolase; Serine protease; Glycoprotein; Signal.

SIGNAL 1 22

CHAIN 23 293

ACT\_SITE 108 108

ACT\_SITE 153 153

ACT\_SITE 245 245

DISULFID 73 206

DISULFID 93 109

DISULFID 178 279

DISULFID 185 251

DISULFID 217 231

DISULFID 241 266

CARBOHYD 69 69

CARBOHYD 173 173

CARBOHYD 208 208

CARBOHYD 252 252

CONFLICT 25 55

SEQUENCE 293 AA; 32020 MW; D92C92F5609E5946 CRC64;

FT ACT SITE 108 108 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT ACT SITE 153 153 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT ACT SITE 245 245 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT DISULFID 73 206 BY SIMILARITY.

FT DISULFID 93 109 BY SIMILARITY.

FT DISULFID 178 279 BY SIMILARITY.

FT DISULFID 185 251 BY SIMILARITY.

FT DISULFID 217 231 BY SIMILARITY.

FT DISULFID 241 266 BY SIMILARITY.

FT CARBOHYD 69 69 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 173 173 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 208 208 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 252 252 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CONFLICT 25 55 MISSING (IN REF. 3).

SEQUENCE 293 AA; 32020 MW; D92C92F5609E5946 CRC64;

Query Match 76.0%; Score 38; DB 1; Length 293;  
Best Local Similarity 66.7%; Pred. NO. 4.4;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

1 GPLVCRGTL 9  
||:| |:  
247 GPVVCNGSL 255

arch completed: March 1, 2004, 17:30:00  
b time : 7 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

M protein - protein search, using sw model

un on: March 1, 2004, 17:20:41 ; Search time 30.3333 Seconds  
(without alignments)  
93.615 Million cell updates/sec

Title: US-09-905-083-80

Effect score: 50

Sequence: 1 GPLVCRGTL 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25.\*

1: sp\_archaea.\*

2: sp\_bacteria.\*

3: sp\_fungi.\*

4: sp\_human.\*

5: sp\_invertebrate.\*

6: sp\_mammal.\*

7: sp\_mhc.\*

8: sp\_organelle.\*

9: sp\_phase.\*

10: sp\_plant.\*

11: sp\_rodent.\*

12: sp\_virus.\*

13: sp\_vertebrate.\*

14: sp\_unclassified.\*

15: sp\_rvirus.\*

16: sp\_bacteriap.\*

17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DS	ID	Description
1	50	100.0	181	4	Q8NFV7	Q8NFV7 homo sapien
2	50	100.0	253	4	Q8NSN9	Q8NSN9 homo sapien
3	45	90.0	251	11	Q9DBQ8	Q9DBQ8 mus musculus
4	41	82.0	87	11	Q9CQ78	Q9CQ78 mus musculus
5	41	82.0	249	11	Q9QYN4	Q9QYN4 mus musculus
6	41	82.0	276	11	Q9QYN3	Q9QYN3 m hippostas
7	41	82.0	612	2	Q83WX7	Q83WX7 streptomyces
8	41	82.0	621	16	Q8G3L1	Q8G3L1 bifidobacter
9	41	82.0	3201	2	Q8P828	Q8P828 micromonospor
10	41	82.0	3546	2	Q8P830	Q8P830 micromonospor
11	40	80.0	220	4	Q8NCW4	Q8NCW4 homo sapien
12	40	80.0	254	6	Q8XSN6	Q8XSN6 sus scrofa
13	40	80.0	257	13	Q8QG86	Q8QG86 bothriopsis in
14	40	80.0	262	16	Q9XAC9	Q9XAC9 streptomyces
15	40	80.0	286	4	Q8ETG8	Q8ETG8 homo sapien
16	40	80.0	619	16	Q8KEH3	Q8KEH3 chlorobium

Q9YBC2 aeropyrum p  
Q9WQ5 paralichthy  
Q9NR68 homo sapien  
Q9R048 mus musculu  
Q9CV76 mus musculu  
Q8N460 homo sapien  
Q86VJ5 homo sapien  
Q9LVE3 mus musculu  
Q8WZB4 homo sapien  
Q8K597 rattus norv  
Q8CGR4 mus musculu  
Q8IWE9 homo sapien  
Q8IUS5 homo sapien  
Q8CGR6 mus musculu  
Q8PPW7 xanthomonas  
Q8P4U1 xanthomonas  
Q8S05 mus musculu  
Q8BPP6 mus musculu  
Q86YZ7 homo sapien  
Q7X7N0 oryza sativ  
Q86YZ8 mus sapien  
Q8OXD5 mus musculu  
Q8K002 mus musculu  
Q8WX98 homo sapien  
Q8CHD0 mus musculu  
Q62540 mus spretus  
Q9CVU2 mus musculu  
Q9D974 mus musculu  
Q9PUF3 bothriops ja

## ALIGNMENTS

### RESULT 1

Q8NFV7  
ID Q8NFV7 PRELIMINARY; PRT; 181 AA.  
AC Q8NFV7;  
DT 01-OCT-2002 (Tremblrel. 22, Created)  
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)  
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)  
DE Kallikrein 7 short variant protein.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC Tissue-Ovarian Carcinoma;  
RA Dong Y., Kaushal A., Clements J.A.;  
RT "Human kallikrein 7 (KJK7) short variant mRNA from ovarian carcinoma."  
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.  
DR EMBL; AF411215; AAN03663.1; -.  
DR GO; GO:0004263; P:chymotrypsin activity; IEA.  
DR GO; GO:0008233; P:peptidase activity; IEA.  
DR GO; GO:0004295; P:trypsin activity; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR InterPro; IPR009003; Cys\_Ser\_Trypsin.  
DR InterPro; IPR001254; Peptidase\_S1.  
DR InterPro; IPR001314; Peptidase\_S1.  
DR Pfam; PF00089; trypsin; 1.  
DR PRINTS; PR00722; CHYMOTRYPSIN.  
DR SMART; SM00020; TRY\_P\_SPC; 1.  
DR PROSITE; PS0240; TRYPsin\_DOM; 1.  
DR PROSITE; PS0135; TRYPsin\_SER; 1.  
KW Hydrolase; Protease; Serine protease.  
SQ SEQUENCE 181 AA; 19887 MW; 86A28A03B90C2D78 CRC64;

Query Match 100.0%; Score 50; DB 4; Length 181;  
Best Local Similarity 100.0%; Pred. No. 0.089;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Y      1 GPLVCRGTL 9
b      135 GPLVCRGTL 143

RESULT 2
D Q8NS9          PRELIMINARY;      PRT;      253 AA.
C Q8NS9;
T 01-OCT-2002 (TrEMBLrel. 22, Created)
T 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
T 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
E Kallikrein 7 (chymotryptic, stratum corneum).
S Homo sapiens (human); Chordata; Craniata; Vertebrata; Euteleostomi;
K Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
M Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
X NCBI_TaxID=9606;
N [1]
P SEQUENCE FROM N.A.
C TISSUE=Skin;
A Strausberg R.;
L Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
C -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
R GO: GO:0004263; F:chymotrypsin activity; IEA.
R GO: GO:0004263; F:peptidase activity; IEA.
R GO: GO:0004295; F:trypsin activity; IEA.
R GO: GO:0006508; P:proteolysis and peptidolysis; IEA.
R InterPro: IPR009003; Cys Ser trypsin.
R InterPro: IPR001254; Peptidase_S1.
R Pfam: PF00089; trypsin; 1.
R PRINTS; PR00722; CHYMOTRYPSIN.
R SMART; SM00020; TRYSPC; 1.
R PROSITE; PS02040; TRYPSIN_DOM; 1.
R PROSITE; PS00134; TRYPSIN_HIS; 1.
R PROSITE; PS00135; TRYPSIN_SER; 1.
W Hydrolyase; Protease; Serine protease.
Q SEQUENCE 253 AA; 27608 MW; 2D68B6A41B22A668 CRC64;

Query Match 100.0%; Score 50; DB 4; Length 253;
Best Local Similarity 100.0%; Pred. No. 0.12;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2y      1 GPLVCRGTL 9
2b      207 GPLVCRGTL 215

RESULT 3
D Q8BQ8          PRELIMINARY;      PRT;      251 AA.
C Q8BQ8;
T 01-JUN-2001 (TrEMBLrel. 17, Created)
T 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
T 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
E 1200016C12Rik protein.
S Mus musculus (mouse).
K Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
M Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
X NCBI_TaxID=10090;
N [1]
P SEQUENCE FROM N.A.
C STRAIN=C57BL/6J; TISSUE=Lung;
R Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
R Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
R Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
R Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
R Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
R Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
R Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
R Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
R Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
R Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
R Gustincich S., Hill D., Hofmann M., Hume D.A., Kasukawa T., Saito R.,
R Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
R Nordone P., Marchionni L., Mashima J., Mazzarelli J., Sakamoto N.,
R Sasaki H., Sato K., Schoenbach C., Seva T., Shibata Y., Storch K.-F.,
R Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
R Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
R Hayashizaki Y.;
R "Functional annotation of a full-length mouse cDNA collection.";
R Nature 409:685-690(2001).
C -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
R EMBL; AK004807; BAB23579.1; -.
R HSP; P00763; IDPO.
R MEROPS; S01.307; -.
R MGD; MGI:1921082; 1200016C12Rik.
R GO: GO:0004263; F:chymotrypsin activity; IEA.
R GO: GO:0008233; F:peptidase activity; IEA.
R GO: GO:0004295; F:trypsin activity; IEA.
R GO: GO:0006508; P:proteolysis and peptidolysis; IEA.
R InterPro: IPR009003; Cys Ser trypsin.
R InterPro: IPR001254; Peptidase_S1.
R Pfam; PF00089; trypsin; 1.
R PRINTS; PR00722; CHYMOTRYPSIN.
R SMART; SM00020; TRYSPC; 1.
R PROSITE; PS02040; TRYPSIN_DOM; 1.
R PROSITE; PS00134; TRYPSIN_HIS; 1.
R PROSITE; PS00135; TRYPSIN_SER; 1.
W Hydrolyase; Protease; Serine protease.
Q SEQUENCE 251 AA; 28153 MW; F4D667F8C0C4A23 CRC64;

Query Match 90.0%; Score 45; DB 11; Length 251;
Best Local Similarity 88.9%; Pred. No. 1.1;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      1 GPLVCRGTL 9
Db      206 GPLVCRGTL 214

RESULT 4
Q8CQ78          PRELIMINARY;      PRT;      87 AA.
ID Q8CQ78;
AC Q8CQ78;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE 2310015I08Rik protein (Fragment).
GN PRS920 OR 2310015I08RIK.
OS Mus musculus (mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Tongue;
RX Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RX Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RX Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RX Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RX Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RX Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RX Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RX Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RX Sasaki H., Sato K., Schoenbach C., Seva T., Shibata Y., Storch K.-F.,
RX Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RX Gustincich S., Hill D., Hofmann M., Hume D.A., Kasukawa T., Saito R.,
RX Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,

```

A Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
A Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
A Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
A Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kotsuki S.,  
A Hayashizaki Y.,  
T Functional annotation of a full-length mouse cDNA collection.;  
L Nature 409:685-690(2001).  
R EMBL; AK019102; BAB31548.1; -;  
R EMBL; AK009451; BAB26297.1; -;  
R HSSP; P00763; IDPO.  
R MEROPS; S01.257; -;  
R MGD; MGI:1929977; Prss20.  
R GO; GO:0005576; C:extracellular; IDA.  
R InterPro; IPR009003; Cys Ser trypsin.  
R InterPro; IPR001254; Peptidase\_S1.  
R Pfam; PF00089; trypsin; 1.  
R SMART; SM00020; Tryp\_Src; 1.  
R PROSITE; PS50240; TRYPSIN\_DOM; 1.  
R PROSITE; PS00135; TRYPSIN\_SER; 1.  
W Hydrolase; Protease; Serine protease.  
T NON\_TER 1  
T SEQUENCE 87 AA; 9549 MW; B988D0CDG62926EAA CRC64;  
Query Match 82.0%; Score 41; DB 11; Length 87;  
Best Local Similarity 77.8%; Pred. NO. 2.4;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
2Y 1 GPLVCRGTL 9  
42 GPLVCGSL 50  
RESULT 5  
9QYX4 PRELIMINARY; PRT; 249 AA.  
C OSQYN4  
C OSQYN3  
Y 01-MAY-2000 (TREMBlrel. 13, Created)  
Y 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
Y 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
DE Hippotastin.  
DE Mus musculus (Mouse).  
DE Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
DE Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
DE NCBI\_TaxID=10090;  
DE [1]  
DE Kawai J., Kojima Y., Konno H., Kouda M., Koya S., Kurihara C.,  
DE Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,  
DE Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai K.,  
DE Sano H., Sasaki D., Shibata K., Shibata Y., Shinagawa A., Shiraki T.,  
DE Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,  
DE Tejima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,  
DE Muramatsu M., Hayashizaki Y.;  
DE Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
DE [2]  
DE SEQUENCE FROM N.A.  
DE TISSUE=Brain;  
DE Yamauchi N., Mitsui S.;  
DE Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
DE [3]  
DE SEQUENCE FROM N.A.  
DE STRAIN=C57BL/6J; TISSUE=Tongue;  
DE Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,  
DE Arakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,  
DE Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hiraoka T., Hori F.,  
DE Imotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,  
DE Kawai J., Kojima Y., Konno H., Kouda M., Koya S., Kurihara C.,  
DE Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,  
DE Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai K.,  
DE Sano H., Sasaki D., Shibata K., Shibata Y., Shinagawa A., Shiraki T.,  
DE Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,  
DE Tejima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,  
DE Muramatsu M., Hayashizaki Y.;  
DE Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
DE [4]  
DE SEQUENCE FROM N.A.  
DE TISSUE=Brain;  
DE Mitsui S., Yamaguchi N.;  
DE "cDNA cloning of a novel brain serine protease, Hippotastin.";  
DE Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
DE EMBL; AB016226; BAA88825.1; -;  
DE HSSP; P00763; IDPO.  
DE MEROPS; S01.257; -;  
DE MGD; MGI:1929977; Prss20.  
DE GO; GO:0005576; C:extracellular; IDA.  
DE InterPro; IPR009003; Cys Ser trypsin.  
DE InterPro; IPR001254; Peptidase\_S1.  
DE InterPro; IPR001314; Peptidase\_S1A.  
DE Pfam; PF00089; trypsin; 1.  
DE PRINTS; PR00722; CHYMOTRYPSIN.  
DE SMART; SM00020; Tryp\_Src; 1.  
DE PROSITE; PS50240; TRYPSIN\_DOM; 1.  
DE PROSITE; PS00134; TRYPSIN\_HIS; 1.  
DE PROSITE; PS00135; TRYPSIN\_SER; 1.  
SQ SEQUENCE 249 AA; 27604 MW; F9FF9CB457D727D5 CRC64;  
Query Match 82.0%; Score 41; DB 11; Length 249;

Best Local Similarity 77.8%; Pred. NO. 6.5;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 1 GPLVCRGTL 9  
DB 204 GPLVCGSL 212  
RESULT 6  
9QYX3 PRELIMINARY; PRT; 276 AA.  
C OSQYN3  
C OSQYN3  
Y 01-MAY-2000 (TREMBlrel. 13, Created)  
Y 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
Y 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
DE Hippotastin prostate type (Adult male tongue cDNA, RIKEN full-length  
DE enriched library, clone:2310015108 product:protease, serine, 20, full  
DE insert sequence) (Adult male tongue cDNA, RIKEN full-length enriched  
DE library, clone:2310040F07 product:protease, serine, 20, full insert  
DE sequence).  
DE PRSS20.  
DE Mus musculus (Mouse).  
DE Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
DE Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
DE NCBI\_TaxID=10090;  
DE [1]  
DE SEQUENCE FROM N.A.  
DE TISSUE=Prostate;  
DE Yamauchi N., Mitsui S.;  
DE Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
DE [2]  
DE SEQUENCE FROM N.A.  
DE TISSUE=Prostate;  
DE MEDLINE=20525460; PubMed=11072088;  
DE Mitsui S., Okui A., Kominami K., Uemura H., Yamaguchi N.;  
DE "cDNA cloning and tissue-specific splicing variants of mouse  
DE hippotastin/TLSP (PRSS20).";  
DE Biochim. Biophys. Acta 1494:206-210(2000).  
DE [3]  
DE SEQUENCE FROM N.A.  
DE STRAIN=C57BL/6J; TISSUE=Tongue;  
DE Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,  
DE Arakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,  
DE Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hiraoka T., Hori F.,  
DE Imotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,  
DE Kawai J., Kojima Y., Konno H., Kouda M., Koya S., Kurihara C.,  
DE Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,  
DE Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai K.,  
DE Sano H., Sasaki D., Shibata K., Shibata Y., Shinagawa A., Shiraki T.,  
DE Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,  
DE Tejima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,  
DE Muramatsu M., Hayashizaki Y.;  
DE Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
DE [4]  
DE SEQUENCE FROM N.A.  
DE STRAIN=C57BL/6J; TISSUE=Tongue;  
DE The FANTOM Consortium,  
DE "Analysis of the mouse transcriptome based on functional annotation of  
DE 60,770 full-length cDNAs.";  
DE Nature 420:563-573(2002).  
DE [5]  
DE SEQUENCE FROM N.A.  
DE STRAIN=C57BL/6J; TISSUE=Tongue;  
DE RIKEN FANTOM Consortium;  
DE "Functional annotation of a full-length mouse cDNA collection.";  
DE Nature 409:685-690(2001).  
DE [6]  
DE SEQUENCE FROM N.A.  
DE STRAIN=C57BL/6J; TISSUE=Tongue;  
DE MEDLINE=99279253; PubMed=10349636;  
DE Carninci P., Hayashizaki Y.;  
DE "High-efficiency full-length cDNA cloning.";  
DE



```

Meth. Enzymol. 303:19-44 (1999).
[7]
RN SEQUENCE FROM N.A.
RP STRAIN=C57BL/6J; PubMed=11042159;
RC MEDLINE=20499374; PubMed=11042159;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
A Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
T "Normalization and subtraction of cap-trapper-selected cDNAs to
T prepare full-length cDNA libraries for rapid discovery of new genes.";
L Genome Res. 10:1617-1630 (2000).
[8]
RN SEQUENCE FROM N.A.
RP STRAIN=C57BL/6J; Tissue=Tongue;
RC MEDLINE=20330913; PubMed=11076861;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
A Kono H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,
A Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
A Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
A Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
A Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
A Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
T "RIKEN integrated sequence analysis (RISA) system-384-format
T sequencing pipeline with 384 multicapillary sequencer.";
L Genome Res. 10:1757-1771 (2000).
[9]
RN -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
RC MEDLINE=90053144; PubMed=9836424;
RA Kinashi H., Fujii S., Hatanai A., Kurokawa T., Shinkawa H.;
A "Physical mapping of the linear plasmid pSLA2-L and localization of
A the eryAI and actI homologs.";
R Biosci. Biotechnol. Biochem. 62:1892-1897 (1998).
RL EMBL; AB088224; BAC76588.1;
DR GO; GO:0046821; C:extrachromosomal DNA; IEA.
DR Hypothetical protein; Plasmid.
KW SEQUENCE 276 AA; 30753 MW; 90BDC03A8AB178D6 CRC64;
SEQUENCE 276 AA; 30753 MW; 90BDC03A8AB178D6 CRC64;

Query Match 82.0%; Score 41; DB 11; Length 276;
Best Local Similarity 77.8%; Pred. No. 7.2;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

2Y 1 GPLVCRGTL 9
| | | | |
Db 231 GPLVCGSL 239
| | | | |

RESULT 7
Q83WX7 PRELIMINARY; PRT; 612 AA.
AC Q83WX7;
DT 01-JUN-2003 (TrEMBLrel. 24; Created)
DT 01-JUN-2003 (TrEMBLrel. 24; Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25; Last annotation update)
DE Hypothetical protein.
DE Streptomyces rochei (Streptomyces parvullus).
OG Plasmid pSLA2-L.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1928;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=7434AN4;
RA Mochizuki S., Hiratsu K., Suwa M., Ishii T., Sugino F., Yamada K.,
RA Kinashi H.;
RT "The large linear plasmid pSLA2-L of Streptomyces rochei has an
RT unusually condensed gene organization for secondary metabolism.";

Meth. Microbiol. 0:0-0 (2003).
[2]
RN SEQUENCE FROM N.A.
RP STRAIN=7434AN4;
RC MEDLINE=20408175; PubMed=10954087;
RA Hiratsu K., Mochizuki S., Kinashi H.;
RT "Cloning and analysis of the replication origin and the telomeres of
RT the large linear plasmid pSLA2-L in Streptomyces rochei.";
L Mol. Gen. Genet. 263:1015-1021 (2000).
[3]
RN SEQUENCE FROM N.A.
RP STRAIN=7434AN4;
RC MEDLINE=20231737; PubMed=10767533;
RA Suwa M., Sugino H., Sasaoka A., Mori E., Fujii S., Shinkawa H.,
RA Nimi O., Kinashi H.;
RT "Identification of two polyketide synthase gene clusters on the linear
RT plasmid pSLA2-L in Streptomyces rochei.";
L Gene 246:123-131 (2000).
[4]
RN SEQUENCE FROM N.A.
RP STRAIN=7434AN4;
RC MEDLINE=99053144; PubMed=9836424;
RA Kinashi H., Fujii S., Hatanai A., Kurokawa T., Shinkawa H.;
A "Physical mapping of the linear plasmid pSLA2-L and localization of
A the eryAI and actI homologs.";
R Biosci. Biotechnol. Biochem. 62:1892-1897 (1998).
RL EMBL; AB088224; BAC76588.1;
DR GO; GO:0046821; C:extrachromosomal DNA; IEA.
DR Hypothetical protein; Plasmid.
KW SEQUENCE 612 AA; 66525 MW; 9BF0E1EB8D3110FC CRC64;
SEQUENCE 612 AA; 66525 MW; 9BF0E1EB8D3110FC CRC64;

Query Match 82.0%; Score 41; DB 2; Length 612;
Best Local Similarity 77.8%; Pred. No. 16;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
| | | | |
Db 250 GPLVCGTL 258
| | | | |

RESULT 8
Q8G3L1 PRELIMINARY; PRT; 621 AA.
AC Q8G3L1;
DT 01-MAR-2003 (TrEMBLrel. 23; Created)
DT 01-MAR-2003 (TrEMBLrel. 23; Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25; Last annotation update)
DE Probable long-chain-fatty-acid-CoA ligase, long-chain acyl-CoA
DE synthetase.
GN FADD4 OR BLI748.
OS Bifidobacterium longum.
OC Bacteria; Actinobacteria; Actinobacteridae; Bifidobacteriales;
OC Bifidobacteriaceae; Bifidobacterium.
OX NCBI_TaxID=216816;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCC 2705;
RA Schell M.A., Karmirantzou M., Snel B., Vilanova D., Berger B.,
RA Pessi G., Zwahlen M.-C., Desiere F., Bork P., Delley M.,
RA Pridmore R.D., Arigoni F.;
RT "The genome sequence of Bifidobacterium longum reflects its adaptation
RT to the human gastrointestinal tract.";
L Proc. Natl. Acad. Sci. U.S.A. 99:14422-14427 (2002).
RL EMBL; AS014808; AN2553.1;
DR GO; GO:0016874; F:ligase activity; IEA.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR002086; AMP-binding.
DR InterPro; IPR000873; AMP-bind.
DR Pfam; PF00501; AMP-binding; 1.
DR PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; 1.
DR PROSITE; PS00455; AMP_BINDING; 1.
```

QW Ligase; Complete proteome.  
 QY SEQUENCE 521 AA; 67948 MW; 5BCDDB4B5BF3083A CRC64;  
 Query Match 82.0%; Score 41; DB 16; Length 621;  
 Best Local Similarity 100.0%; Pred. No. 16;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2Y 1 GPLVCRG 7  
 |||||  
 436 GPLVCRG 442

2B RESULT 9  
 29F828 PRELIMINARY; PRT; 3201 AA.  
 ID AC Q9F828  
 DT 01-MAR-2001 (TREMELrel. 16, Created)  
 DT 01-MAR-2001 (TREMELrel. 16, Last sequence update)  
 DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)  
 DE Megalomicin 6-deoxyerythronolide B synthase 3.  
 GN MEGALIII.  
 OS Micromonospora megalomicea subsp. nigra.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Micromonosporineae; Micromonosporaceae; Micromonospora.  
 OX NCBI\_TaxID=136926;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NRRL3275;  
 RX MEDLINE=20430101; PubMed=10972798;  
 RA Volchegursky Y., Hu Z., Katz L., McDaniel R.;  
 RT "Biosynthesis of the anti-parasitic agent megalomicin: transformation  
 of erythromycin to megalomicin in *Saccharopolyspora erythraea*.";  
 RL Mol. Microbiol. 37:752-762(2000).  
 CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES  
 (SDR) FAMILY.  
 CC EMBL; AF263245; AAG13919.1; -.  
 DR HSP; 25715; IMLA.  
 DR GO; GO:0016798; F:hydrolase activity, acting on ester bonds; IEA.  
 DR GO; GO:0016491; F:oxidoreductase activity; IEA.  
 DR GO; GO:0016740; F:transferase activity; IEA.  
 DR GO; GO:0016740; F:transferase activity; IEA.  
 DR GO; GO:0005215; P:transporter activity; IEA.  
 DR GO; GO:0006633; P:fatty acid biosynthesis; IEA.  
 DR GO; GO:0008152; P:metabolism; IEA.  
 DR GO; GO:0008810; P:transport; IEA.  
 DR InterPro; IPR001227; AC trans.  
 DR InterPro; IPR002198; ADH short.  
 DR InterPro; IPR000794; Ketoacyl synth.  
 DR InterPro; IPR000566; Lipocin cytFABP.  
 DR InterPro; IPR006162; Ppantne\_S.  
 DR InterPro; IPR006163; Pp\_bind\_3.  
 DR Pfam; PF00698; Acyl\_transf; 3.  
 DR Pfam; PF00106; adh\_short; 2.  
 DR Pfam; PF00109; ketoacyl-synt; 2.  
 DR Pfam; PF02801; ketoacyl-synt\_C; 2.  
 DR Pfam; PF00550; pp-binding; 3.  
 DR PROSITE; PS00075; ACP DOMAIN; 3.  
 DR PROSITE; PS00606; B\_KETOACYL SYNTHASE; 2.  
 DR PROSITE; PS00213; LIPOCALIN\_2.  
 DR PROSITE; PS00012; PHOSPHOPANTHEINE; 2.  
 KW Oxidoreductase; Phosphopantetheine; Transferase.  
 SQ SEQUENCE 3201 AA; 334781 MW; 44BDA30E14855650 CRC64;

Query Match 82.0%; Score 41; DB 2; Length 3201;  
 Best Local Similarity 77.8%; Pred. No. 77;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9  
 |||||  
 1877 GPMVCRGGL 1885

2B RESULT 10  
 Q9F830 PRELIMINARY; PRT; 3546 AA.  
 ID AC Q9F830  
 DT 01-MAR-2001 (TREMELrel. 16, Created)  
 DT 01-MAR-2001 (TREMELrel. 16, Last sequence update)  
 DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)  
 DE Megalomicin 6-deoxyerythronolide B synthase 1.  
 GN MEGAL.  
 OS Micromonospora megalomicea subsp. nigra.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Micromonosporineae; Micromonosporaceae; Micromonospora.  
 OX NCBI\_TaxID=136926;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NRRL3275;  
 RX MEDLINE=20430101; PubMed=10972798;  
 RA Volchegursky Y., Hu Z., Katz L., McDaniel R.;  
 RT "Biosynthesis of the anti-parasitic agent megalomicin: transformation  
 of erythromycin to megalomicin in *Saccharopolyspora erythraea*.";  
 RL Mol. Microbiol. 37:752-762(2000).  
 CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES  
 (SDR) FAMILY.  
 CC EMBL; AF263245; AAG13917.1; -.  
 DR HSP; 25715; IMLA.  
 DR GO; GO:0016491; F:oxidoreductase activity; IEA.  
 DR GO; GO:0016740; F:transferase activity; IEA.  
 DR GO; GO:0005215; P:transporter activity; IEA.  
 DR GO; GO:0006633; P:fatty acid biosynthesis; IEA.  
 DR GO; GO:0008152; P:metabolism; IEA.  
 DR GO; GO:0008810; P:transport; IEA.  
 DR InterPro; IPR001227; AC trans.  
 DR InterPro; IPR002198; ADH short.  
 DR InterPro; IPR000794; Ketoacyl synth.  
 DR InterPro; IPR000566; Lipocin cytFABP.  
 DR InterPro; IPR006162; Ppantne\_S.  
 DR InterPro; IPR006163; Pp\_bind\_3.  
 DR Pfam; PF00698; Acyl\_transf; 3.  
 DR Pfam; PF00106; adh\_short; 2.  
 DR Pfam; PF00109; ketoacyl-synt; 2.  
 DR Pfam; PF02801; ketoacyl-synt\_C; 2.  
 DR Pfam; PF00550; pp-binding; 3.  
 DR PROSITE; PS00075; ACP DOMAIN; 3.  
 DR PROSITE; PS00606; B\_KETOACYL SYNTHASE; 2.  
 DR PROSITE; PS00012; PHOSPHOPANTHEINE; 2.  
 KW Oxidoreductase; Phosphopantetheine; Transferase.  
 SQ SEQUENCE 3546 AA; 371191 MW; 86C6794E95415B8C CRC64;

Query Match 82.0%; Score 41; DB 2; Length 3546;  
 Best Local Similarity 77.8%; Pred. No. 85;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9  
 |||||  
 2397 GPMVCRGGL 2405

2B RESULT 11  
 Q9NCW4 PRELIMINARY; PRT; 220 AA.  
 ID AC Q9NCW4  
 DT 01-OCT-2002 (TREMELrel. 22, Created)  
 DT 01-OCT-2002 (TREMELrel. 22, Last sequence update)  
 DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)  
 DE Prostate specific antigen precursor.  
 GN KLK3.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Prostate;  
 RA Heuze-Vourc'h N., Courty Y.;

\*Complex alternative splicing of the hKLK3 gene coding for the tumour marker PSA (prostate-specific-antigen).";  
Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.

1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.

EMBL: AJ459782; CAD30844.1; -.

GO: GO:0004263; F:chymotrypsin activity; IEA.

GO: GO:0008233; F:peptidase activity; IEA.

GO: GO:0004295; F:trypsin activity; IEA.

GO: GO:0006508; P:proteolysis and peptidolysis; IEA.

InterPro: IPR009003; Cys\_Ser\_trypsin.

InterPro: IPR001254; Peptidase S1.

InterPro: IPR001314; Peptidase\_SIA.

Pfam: PF00089; trypsin.1

PRINTS: PR00722; CHYMOTRYPSIN.

SMART: SM00020; Tryp\_SPC; 1.

PROSITE: PS0240; TRYPsin DOM; 1.

PROSITE: PS00134; TRYPsin HIS; 1.

PROSITE: PS00135; TRYPsin SER; 1.

Hydrolase; Protease; Serine protease; Signal.

SIGNAL 1 17 POTENTIAL.

CHAIN 25 220 PROSTATE SPECIFIC ANTIGEN.

SEQUENCE 220 AA; 24341 MW; 15C886A50A248D CRC64;

Query Match 80.0%; Score 40; DB 4; Length 220;

Best Local Similarity 77.8%; Pred. No. 9;

Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

2y 1 GPLVCRGTL 9

174 GPLVCNGVL 182

RESULT 12

29XSN6

ID Q9XSN6 PRELIMINARY; PRT; 254 AA.

AC Q9XSN6;

DT 01-NOV-1999 (TrEMBLrel. 12, Created)

DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Enamel matrix serine proteinase 1 precursor.

OS Sus scrofa (Pig).

OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.

OC NCBI\_TaxID=9823;

[1]

SEQUENCE FROM N.A.

RX MEDLINE=98126310; PubMed=9465170;

RA Simmer J.P., Fukae M., Tanabe T., Yamakoshi Y., Uchida T., Xue J.,

RA Margolis H.C., Shimizu M., DeHart B.C., Hu C.-C., Bartlett J.D.,

RT "Purification, characterization, and cloning of enamel matrix serine

proteinase 1."; Res. 77:377-386(1998).

RL J. Dent. Res. 77:377-386(1998).

CC 1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.

EMBL: U76256; AAB94638.1; -.

DR HSSP; P00763; LDPO.

DR MEROPS; S01.251; -.

DR GO: GO:0004263; F:chymotrypsin activity; IEA.

DR GO: GO:0008233; F:peptidase activity; IEA.

DR GO: GO:0004295; F:trypsin activity; IEA.

DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.

InterPro: IPR009003; Cys\_Ser\_trypsin.

InterPro: IPR001254; Peptidase S1.

InterPro: IPR001314; Peptidase\_SIA.

Pfam: PF00089; trypsin.1

PRINTS: PR00722; CHYMOTRYPSIN.

SMART: SM00020; Tryp\_SPC; 1.

PROSITE: PS0240; TRYPsin DOM; 1.

PROSITE: PS00134; TRYPsin HIS; 1.

PROSITE: PS00135; TRYPsin SER; 1.

Hydrolase; Protease; Serine protease; Signal.

SIGNAL 1 24 POTENTIAL.

CHAIN 31 254 ENAMEL MATRIX SERINE PROTEINASE 1.

SEQUENCE 254 AA; 27235 MW; FD40EF85664406F1 CRC64;

Query Match 80.0%; Score 40; DB 6; Length 254;

Best Local Similarity 66.7%; Pred. No. 10;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY

1 GPLVCRGTL 9

209 GPLCNGSL 217

RESULT 13

Q8QG86

ID Q8QG86 PRELIMINARY; PRT; 257 AA.

AC Q8QG86;

DT 01-JUN-2002 (TrEMBLrel. 21, Created)

DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Serine proteinase.

OS Bothrops insularis (Island jararaca) (Quelmada jararaca).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidae;

OC Viperidae; Crotalinae; Bothrops.

OC NCBI\_TaxID=8723;

[1]

SEQUENCE FROM N.A.

RC TISSUE=Venom gland;

RA Junqueira-de-Azevedo I.L.M., Ho P.L.;

RT "A survey of gene expression and diversity in the venom glands of the

RT pitviper Bothrops insularis through the generation of Expressed

RT Sequence Tags (ESTs).";

RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF490536; AAM09695.1; -.

DR HSSP; P00761; IAN1.

DR GO: GO:0004263; F:chymotrypsin activity; IEA.

DR GO: GO:0008233; F:peptidase activity; IEA.

DR GO: GO:0004295; F:trypsin activity; IEA.

DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.

InterPro: IPR009003; Cys\_Ser\_trypsin.

InterPro: IPR001254; Peptidase\_S1.

InterPro: IPR001314; Peptidase\_SIA.

Pfam: PF00089; trypsin.1

PRINTS: PR00722; CHYMOTRYPSIN.

SMART: SM00020; Tryp\_SPC; 1.

PROSITE: PS0240; TRYPsin DOM; 1.

PROSITE: PS00134; TRYPsin HIS; 1.

PROSITE: PS00135; TRYPsin SER; 1.

Hydrolase; Protease; Serine protease.

SEQUENCE 257 AA; 28357 MW; BFFBCE3C4861C967 CRC64;

Query Match 80.0%; Score 40; DB 13; Length 257;

Best Local Similarity 75.0%; Pred. No. 10;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY

1 GPLVCRGT 8

205 GPLCNGT 212

RESULT 14

Q9XAC9

ID Q9XAC9 PRELIMINARY; PRT; 262 AA.

AC Q9XAC9;

DT 01-NOV-1999 (TrEMBLrel. 12, Created)

DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)

DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)

DE Putative aminoglycoside acetyltransferase.

GN SCO1927 OR SCC22.09.

OS Streptomyces coelicolor.

OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

OC Streptomycineae; Streptomycetaceae; Streptomycetes.

OC NCBI\_TaxID=1902;

[1]

SEQUENCE FROM N.A.

IC STRAIN=A3(2);  
LA Seeger K., Harris D.;  
UN Submitted (JUL-1999) to the EMBL/GenBank/DBSJ databases.  
[2]  
UN SEQUENCE FROM N.A.  
IC STRAIN=A3(2);  
LA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;  
UN Submitted (JUL-1999) to the EMBL/GenBank/DBSJ databases.  
[3]  
UN SEQUENCE FROM N.A.  
IC STRAIN=A3(2);  
UN MEDLINE=97000351; PubMed=8843436;  
LA Radenbach M., Kieser H.M., Denapante D., Eichner A., Cullum J.,  
LA Kinashi H., Hopwood D.A.;  
UN "A set of ordered cosmids and a detailed genetic and physical map for  
UN the 8 Mb streptomyces coelicolor A3(2) chromosome.";  
UN Mol. Microbiol. 21:77-96(1996).  
[4]  
UN SEQUENCE FROM N.A.  
IC STRAIN=A3(2) / M145;  
UN MEDLINE=21996410; PubMed=12000953;  
LA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,  
LA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,  
LA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
LA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,  
LA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,  
LA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,  
LA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,  
LA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,  
LA Hopwood D.A.;  
UN "Complete genome sequence of the model actinomycete Streptomyces  
UN coelicolor A3(2).";  
UN Nature 417:141-147(2002).  
UN EMBL: AL939110; CAB50752.1; -.  
UN PIR: T35999; T35999.  
UN GO: GO:0016991; F:gentamicin 3'-N-acetyltransferase activity; IEA.  
UN GO: GO:0016740; F:transferase activity; IEA.  
UN InterPro: IPR003679; Antibiotic NAT.  
UN Pfam: PF02522; Antibiotic NAT; I.  
UN Transferase; Complete proteome.  
UN SEQUENCE 262 AA; 27930 MW; 0D239D41795A0B93 CRC64;

Query Match 80.0%; Score 40; DB 16; Length 262;  
Best Local Similarity 88.9%; Pred. No. 11;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GPLVCRGTL 9  
Db 8 GPLVTRGTL 16  
|||||

RESULT 15  
Q86TGS PRELIMINARY; PRT; 286 AA.  
AC Q86TGS;  
DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Kallikrein 3, (Prostate specific antigen) (Fragment).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Peripheral Nervous System;  
RA Strausberg R.;  
RL Submitted (APR-2003) to the EMBL/GenBank/DBSJ databases.  
DR EMBL: BC050595; AAH50595.1; -.  
DR GO: GO:0004263; F:chymotrypsin activity; IEA.  
DR GO: GO:0004295; F:trypsin activity; IEA.  
DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR InterPro: IPR009003; Cys\_Ser\_trypsin.

DR InterPro: IPR001254; Peptidase\_S1.  
DR InterPro: IPR001314; Peptidase\_S1A.  
DR Pfam: PF00089; trypsin; 1.  
DR PRINTS: PRO0722; CHYMOTRYPSIN.  
DR SMART: SMC0020; TRY\_SPC; 1.  
DR PROSITE: PS02040; TRYPSIN\_DOM; 1.  
DR PROSITE: PS00134; TRYPSIN\_HIS; 1.  
DR PROSITE: PS00135; TRYPSIN\_SER; 1.  
FT NON\_TER 1  
SQ SEQUENCE 286 AA; 31665 MW; 711E2CE3ACE4FED CRC64;  
  
Query Match 80.0%; Score 40; DB 4; Length 286;  
Best Local Similarity 77.8%; Pred. No. 12;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
Qy 1 GPLVCRGTL 9  
Db 240 GPLVCGVGL 248  
|||||

Search completed: March 1, 2004, 17:34:52  
Job time : 32.3333 secs